curves to pulsar powering. Our hybrid suggestion is that the peak luminosity is related to Ni^{56} decay and only the long-term decline is pulsar-powered.

Alternatively, a slow decline could be due to other effects (interaction of ejecta with unusually dense surroundings, for example); in that case, one must ask what the long-term behavior of type II's is, because eventually pulsar activation should become apparent just as it has for the Crab nebula. Branch and Cowan (*24*) have, for example, detected a radio source as bright as Cas A only 26 years after SN 1961v (*25*). Observationally, there is the selection effect that most of these events are observed in distant galaxies and can only be followed down for a limited period. The expectation that the remnant will continue to fade may also bias observational effort after about 6 magnitudes of fading (*19, 25, 26*). For a typical type II supernova, excitation by a central pulsar would not be expected after 6 magnitudes of decay unless the spin-down power exceeded 5×10^{39} ergs/sec. Because SN 1987a is starting off a factor 50 to 100 times dimmer, it could reach the pulsar luminosity threshold much sooner, yet 7 magnitudes of decline would be required even before SN 1987a could be powered by a pulsar of typical 10^{12} Gauss magnetic field and entirely plausible 20-msec period. Thanks to its proximity, SN 1987a should be observable over a much larger magnitude range. A change of 20 magnitudes would take at least 4.0 years at the Co^{56} decay rate and flattening at this point would indicate powering by even a common 0.24-second pulsar. A dramatic leveling off of the light curve is therefore expected within a few years, possibly as soon as 10 months after birth of SN 1987a.

The illustrative estimates presented here can obviously be adjusted to accommodate different initial pulsar spin and magnetic field strengths, and different masses and expansion velocities of the inner heavy element shell and outer hydrogen-helium shell. Given the richness of these parametric dependences, it is possible that some of the variability of the type II supernovae light curves may have a natural explanation in terms of pulsar powering, at least at late times. In summary, we suggest that the existence of a Crab-like pulsar in SN 1987a would make itself known through detection of unpulsed (and possibly pulsed) hard gamma rays within about a year, accompanied by a deviation of the optical light curve from that predicted from Co^{56} decay (*27*). An early interpretation pointing to a pulsar is the possible excitation of nearby matter by a jet from the remnant (*28*), which would indicate a high pressure interior breaking out or intrinsic jet formation (*29*).

REFERENCES AND NOTES

1. J. H. Bahcall, M. J. Rees, E. E. Salpeter, *Astrophys. J* **162**, 737 (1970).
2. J. P. Ostriker, *Nature (London)* **327**, 287 (1987).
3. S. E. Woolsley, P. A. Pinto, L. Ensman, in preparation.
4. R. N. Manchester and J. H. Taylor, *Pulsars* (Freeman, New York, 1977).
5. R. D. Blandford, J. H. Applegate, L. Hernquist, *Mon. Not. R. Astron. Soc.* **204**, 1025 (1983).
6. R. M. Bionta *et al.*, *Phys. Rev. Lett.* **58**, 1494 (1987).
7. K. Hirata *et al.*, *ibid.*, p. 1490.
8. M. A. Ruderman, *Annu. Rev. Astron Astrophys.* **10**, 427 (1972).
9. M. J. Rees and J. E. Gunn, *Mon. Not. R. Astron.* **167**, 1 (1974).
10. C. F. Kennel and F. V. Coroniti, *Astrophys. J.* **283**, 694 (1984).
11. ——, *ibid.*, p. 710.
12. These expectations are: (i) a pulsar wind confined by the remnant even after 1000 years, (ii) a positronic wind with Lorentz factors of about 3×10^6 that may be relatively insensitive to rotation rate and magnetic field because it is determined by pair production physics [M. A. Ruderman and P. G. Sutherland, *Astrophys. J.* **196**, 51 (1975)], (iii) the wind is shocked at about 0.1 of the inner radius if the confining remnant, and (iv) synchrotron radiation amounts to perhaps 20% of the total output.
13. G. W. Grodstein, *National Bureau of Standards Circular No. 583* (U.S. Government Printing Office, Washington, DC, 1957), p. 23.
14. One can estimate the latest that dust formation can be initiated by equating the accretion rate $n\sigma V_{th}$ of a dust particle (σ being the cross-sectional area, n the gas concentration, and V_{th} the thermal velocities of the gas particles) with mean rate would take to build up the dust particle of N atoms, N/t, at that epoch. Taking $N = 10^{12}$, $\sigma = 10^{-8}$ cm^2, $V_{th} = 5 \times 10^4$ cm/sec (thermal velocity at a condensation temperature of 10^3 degrees) and $n = 1.2 \times 10^{30}/t$ sec^3 cm^{-3} for the density of heavy elements then gives $t_{crit} \approx 280$ days, too early for full-scale condensation. If instead only about 10^{-3} by weight of micrometer-sized dust is formed (more typical of the intersatellar medium), unit optical depth corresponds to about 0.3 g/cm^2, essentially the Thomson optical depth.
15. J. Middleditch, C. Pennypacker, M. S. Burns, *Astrophys. J.* **273**, 261 (1983).
16. S. A. Colgate and C. McKee, *ibid.* **157**, 623 (1969).
17. W. Baade, *ibid.* **102**, 309 (1945).
18. P. G. Sutherland and J. C. Wheeler, *ibid.* **280**, 282 (1984).
19. R. Barbon, F. Ciatti, L. Rosino, *Astron. Astrophys.* 72, 287 (1979).
20. J. P. Ostriker and J. E. Gunn, *Astrophys. J.* **164**, L95 (1971).
21. P. Bodenheimer and J. P. Ostriker, *ibid.* **191**, 465 (1974).
22. B. Gaffet, *ibid.* **216**, 565 (1977).
23. ——, *ibid.*, p. 852.
24. D. Branch and J. J. Cowan, *ibid.* **297**, L33 (1985).
25. The "slow" supernova SN 1961v in the 12-megaparsec-distant galaxy NGC 1058 was a type II and was studied over 5 or 6 years as it decayed from apparent magnitude 12 down to magnitude 21 [D. Branch and J. L. Greenstein, *Astrophys. J.* **167**, 89 (1971); R. A. Fesen, *ibid.* **297**, L31 (1985)] before it faded from view, declining with a long half-life of about 450 days [V. P. Utrubin, *Astrophys. Space Sci.* **98**, 115 (1984)]. Also, SN 1987a was not the first supernova to have an identified progenitor. There was a 18th magnitude object at the site of SN 1961v before the outburst. At the inferred distance, such a bright object could not have simply been a red or blue giant.
26. R. P. Kirshner and A. Uomoto [*Astrophys. J.* **308**, 685 (1986)] have followed a type II supernova for 2 years and find that the line emission in fact follows the Co^{56} decay rate quite closely, with the continuum emission, having the departure from this simple behavior, again suggestive of a separate and comparable energy source.
27. T. Shigeyama, N. Nomoto, M. Hashimoto, D. Sugimoto, *Nature (London)* **238**, 320 (1987).
28. M. J. Rees, *ibid.*, p. 207.
29. F. C. Michel *Astrophys. J.* **288**, 138 (1985).

30. We appreciate valuable discussions with A. Porter, J. Greenstein, R. Romani, S. Phinney, and R. Blandford. For one of us (F.C.M.) this research was supported by the National Science Foundation under AST-8511709 and by the Kellogg Radiation Laboratory under PHY-8505682 and 8604197. C.F.K. would like to thank the California Institute of Technology for its hospitality and the National Aeronautics and Space Administration, through NGL-05-007-190, for its support.

10 August 1987; accepted 2 October 1987

A Substantial Bias in Nonparametric Tests for Periodicity in Geophysical Data

STEPHEN M. STIGLER AND MELISSA J. WAGNER

A nonparametric test that has been used to conclude that extinction rates are periodic with a period of 26 million years is shown to be substantially biased toward this conclusion, regardless of whether or not the data are periodic in origin (and, indeed, regardless of the actual period if they are in fact periodic). The test is shown to be sensitive to measurement error of a type expected with these data (early recording of extinctions due to missing fossil specimens, the "Signor-Lipps effect"), and it is shown that because of the unequal spacing in time, such models may be expected to produce statistically significant but artifactual periods of (in this case) exactly 26 million years over the span of time actually used.

THERE HAS BEEN CONSIDERABLE INTEREST in recent years in the analysis of time series data in the geophysical sciences, with the goal of determining whether patterns, particularly periodic patterns, can be detected that depart significantly from what might be expected from a

Department of Statistics, University of Chicago, Chicago, IL 60637.

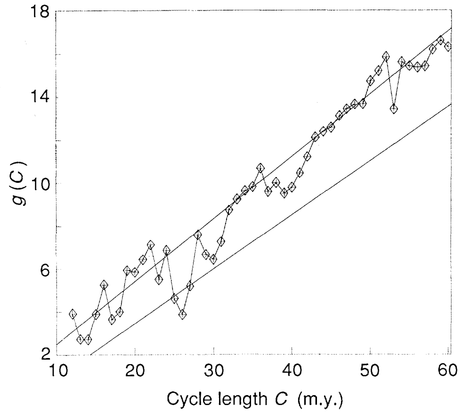

Fig. 1. Values of the measure of fit $g(C)$ (jagged line) computed from the family extinction data (*1*) for C from 12 through 60. Also shown are least-squares fits to Monte Carlo estimates of the mean of $g(C)$ for random series (upper line) and the lower 1% points of the distribution of $g(C)$ for each C for random series (lower line). The measure of fit $d(C)$ is found by subtracting the lower line from $g(C)$, and for the family extinction data the best fitting periodic cycle length as determined by the minimum value of $d(C)$ is $D = -1.05$ achieved at $C = 26$ m.y.

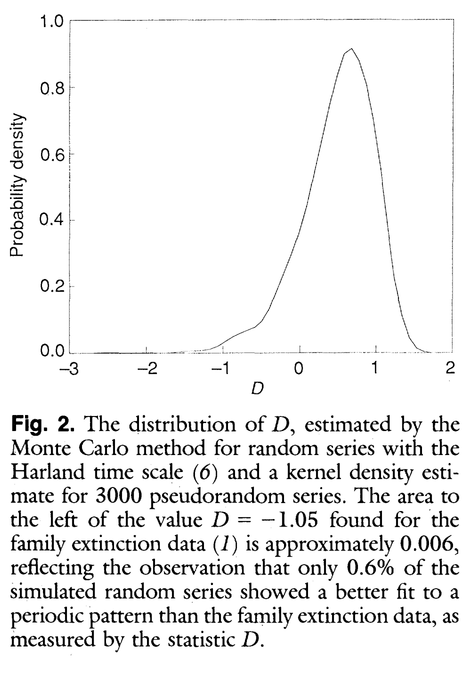

Fig. 2. The distribution of D, estimated by the Monte Carlo method for random series with the Harland time scale (*6*) and a kernel density estimate for 3000 pseudorandom series. The area to the left of the value $D = -1.05$ found for the family extinction data (*1*) is approximately 0.006, reflecting the observation that only 0.6% of the simulated random series showed a better fit to a periodic pattern than the family extinction data, as measured by the statistic D.

series that is in some sense random. For example, Raup and Sepkoski (*1, 2*) examined records of extinctions over the past 250 million years (m.y.) and found evidence of a period of 26 m.y. in the peaks of extinctions, and Raup (*3*) and Stothers (*4*) investigated records of magnetic reversals over the past 165 m.y. with an eye toward determining whether or not a periodic component in variations in their rate could be discovered. In these and other examples (*5*), a simulation-based nonparametric test was employed. The temptation to use such a test is strong, since there are characteristics of the problem and the data that may make standard time-series techniques inappropriate. For one, stratigraphic epochs are of unequal duration, and hence the observed rates are not equally spaced in time. In addition, it is desirable that the test be sensitive to a wide range of different periodic patterns, including both smooth (for example, a sine wave) and sharply punctuated patterns.

The test most frequently employed in the studies cited above is one that was suggested by Stothers (*5*) for investigating the solar activity cycle and was later and independently proposed by Raup and Sepkoski (*1*) for use with extinction data. The versions employed differ in nonessential ways; the version described here is equivalent to that of Raup and Sepkoski, although the notation differs from theirs. The data are supposed to take the form of a series of rates or intensities $R_1, R_2, \ldots, R_n$ observed at or identified with time points $t_1, t_2, \ldots, t_n$. For example, the rate R_i may be the number of extinctions observed to fall in the epoch from t_{i-1} to t_i divided by the length of the epoch, $t_i - t_{i-1}$. The time points need not be equally spaced.

First, the series is examined for peaks: there is a peak at t_i if $R_i > R_{i-1}$ and $R_i > R_{i+1}$. Let $P_1 < P_2 < \ldots < P_k$ be the time points at which there are peaks. For each of a possible range of cycles of duration C m.y. (say, for $12 \le C \le 60$, as with the extinction data), a grid of time points equally spaced by C is adjusted to give the closest agreement to the observed peaks, in the sense that a measure of fit $g(C)$ is minimized. That is, for each t_0 and each P_i, the time point of the form

$$\hat{P}_i = t_0 + m_i C$$

(where m_i is an integer) is found that is closest to P_i. Then t_0 is adjusted and the process is iterated to minimize the standard deviation of the errors $e_i = P_i - \hat{P}_i$

$$g(C) = \left[\frac{1}{k-1}\sum_{i=1}^{k}(e_i - \bar{e})^2\right]^{1/2}$$

The minimized $g(C)$ will of course have $\bar{e} = 0$. The measure of fit $g(C)$ is then examined to determine whether it is sufficiently small for some C to reject the hypothesis of a random series. Figure 1 [a recomputation of part of figure 2 of Raup and Sepkoski (*1*)] shows $g(C)$ for the data on extinction of families; the dip at $C = 26$ to $g(C) = 3.85$ was judged by them to be significant since only a small fraction of simulated series of random rates produced a dip this low at $C = 26$. Subsequent analyses (*2*) that used minor variations on the data and took account of the fact that many potentially significant cycles were being evaluated simultaneously raised the estimated level of significance somewhat, but the deviation was still judged significant at the 1% level, a result in accordance with our analysis below.

Nonparametric tests are appealing in that few assumptions are necessary for their validity. But that advantage comes at a cost: they are often not able to detect effects that it would be desirable to detect (that is, they may lack power), and they may be sensitive to departures from the null hypothesis that are not of interest. These potential difficulties are compounded by the fact that the test statistic is frequently not easily amenable to analytic study, and it can require extensive simulation merely to begin to assess the potential magnitude of these costs. The present study was motivated by a desire both to evaluate the performance characteristics of this particular test (and thus gain a better understanding of the implications for the study of data on extinction rates) and to illustrate in general the kinds of pitfalls that can be encountered in applying nonparametric procedures in complex settings.

The time scale. The evaluation of the test must depend on the time scale, that is, on the total time span covered and on the intervals between time points and their pattern. Most of this study was performed with the use of 40 time points from the Harland time scale (*6*), ranging from 253 m.y. before present (B.P.) (in the Late Permian) to 5.1 m.y. B.P. (at the end of the Miocene). This choice was to provide comparability with Raup and Sepkoski (*1, 2*); however, where

Table 1. Estimated levels of significance (P values) for the nonparametric test for periodicity in the family extinction data under various hypothetical models. The test is not particularly sensitive to autoregressive models, but it is sensitive to moving average models and strongly periodic models. A period in the range of 25 to 30 m.y. was only reliably detected here, however, when the signal-to-noise ratio θ exceeded 2.

Model	θ	Estimated $P(D < -1.05)$
Autoregressive	−0.1	0.002
	0.1	0.010
	0.2	0.016
	0.3	0.025
	0.4	0.036
Moving average	0.25	0.030
	0.50	0.060
	1.00	0.170
Period P = 26 m.y.	0.25	0.026
	0.50	0.079
	1.00	0.465
	2.50	1.0
Period P = 30 m.y.	0.25	0.018
	0.50	0.030
	1.00	0.188
	2.50	1.0
All models	0.0	0.006

they stopped at the end of the middle Miocene (11.3 m.y. B.P.) we went one time point further, to 5.1 m.y. This was because they judged 11.3 to correspond to a peak extinction rate, and unless we extended the time scale beyond 11.3 it would have been impossible for any of our simulated series to peak at that value. Following Raup and Sepkoski (*1*), we amalgamated the Bajocian and Aalenian into one stage (and thus their common boundary, 181 m.y. B.P., does no' appear as a time point in this study).

Although the results we report depend in important ways on the time scale, minor changes do not have a great effect. In particular, our results are not greatly changed if another closely related version of the time scale for the same period, such as that of Odin (*7*) or Snelling (*8*), is used. Despite this general insensitivity to minor alterations in the scale, it will be inappropriate to evaluate significance by a simulation that shuffles the time scale, as is sometimes done [for example, in some of the studies reported (*1*)]. In the terminology of statistical theory, we are treating the time scale as ancillary to the periodic structure, and all tests are conditional on the scale actually observed. Otherwise, extraneous noise may mask a real effect, or fortuitous patterns in the time scale may make an artifactual effect appear to be real.

The problem of multiple comparisons. In assessing the significance of low values of $g(C)$, care must be taken to guard against exaggeration due to the fact that many simultaneous tests are being performed. Since the data were inspected for low $g(C)$ at any period C (not just at $C = 26$), we are consequently interested in determining the probability that chance alone would produce a deviation as extreme as that we observed at any cycle length. Inspection of Fig. 1 suggests that we must correct for the tendency of $g(C)$ to increase with C in judging deviations. A simulation study was performed to estimate the 1% lower percentage point for $g(C)$ for random series, which was found to increase approximately linearly with C. A least-squares fit gave the equation of the line as $0.25C - 1.60$. [The mean of $g(C)$ was estimated to increase as $0.29C - 0.43$; both these lines are in accord with the slopes shown in figure 2 of Raup and Sepkoski (*1*) or figure 4 of Sepkoski and Raup (*9*).] Hence, in all studies reported here deviations were given in terms of $d(C) = g(C) - 0.25C + 1.60$, and the departure from randomness was judged in terms of D—the minimum value of $d(C)$ for C in the range $12 \leq C \leq 60$—with low values of D indicating a good fit to a cycle of the corresponding length (*10*).

Assessing the significance of a deviation. For

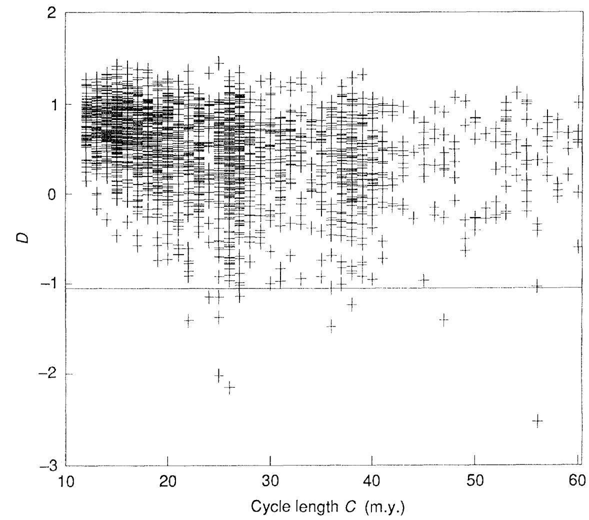

Fig. 3. A plot of D [the minimum value of $d(C)$] versus C (the value of the cycle length for which the best fit to periodicity was found), for 3000 pseudorandom series, with the Harland time scale (*6*). Points below the horizontal line at $D = -1.05$ represent simulated series that showed greater evidence of periodicity than the family extinction data (*1*).

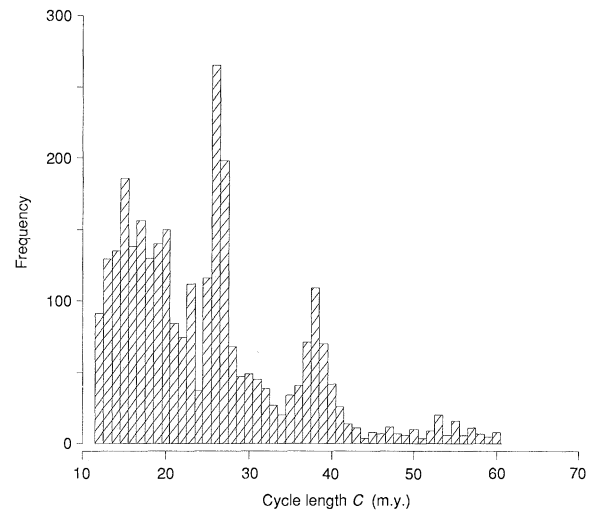

Fig. 4. The frequency distribution of the best fitting cycle length C for 3000 pseudorandom series with the Harland time scale (*6*). Note the pronounced peak at 26 m.y., despite the absence of any periodic component in the simulated extinction rates, which reflects the pattern of the three or four stratigraphic stages of longest duration in that time scale.

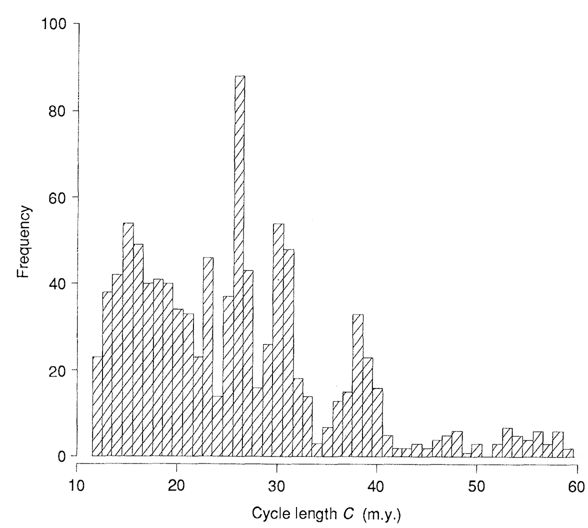

Fig. 5. Frequency distribution generated in the same manner as that in Fig. 4, except that it is based on 1000 series of simulated extinction rates with the Harland time scale (*6*) with a periodic component with cycle length 30 m.y. and with a signal-to-noise ratio of 1:4. Note the pronounced peak at 26 m.y., indicating that such series are more likely to show a best fit to a time cycle of 26 m.y. than to one of 30 m.y.

the data on family extinction, the minimum deviation occurs at 26 m.y.; its value is $d(26) = -1.05$. To determine whether $D = \min d(C) = -1.05$ is a significant departure from what would be expected with random series, a simulation experiment was performed. A total of 3000 series was generated, each series consisting of 40 independent pseudorandom variables U_i (taken as uniformly distributed between 0 and 1), and D was calculated for each series. That is, for each series, we (i) set the 40 rates R_i equal to 40 of the variables U_i, (ii) found as peaks those R_i for which $R_i > R_{i-1}$ and

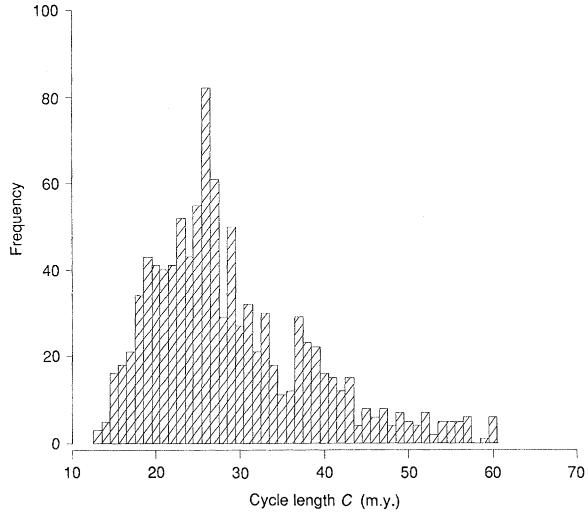

Fig. 6. Frequency distribution generated in the same manner as that in Fig. 4, except that it is based on 1000 series of simulated extinction rates with the Harland time scale (*6*) generated according to a moving average process with θ = 1.0. This process is intended to reflect the effect of measurement error of a type anticipated with such data, due either to the early recording of extinction events because of missing fossil records or to the fractional allocation of events to several stages because of a lack of resolution of the time of the event. Note the pronounced peak at 26 m.y. despite the absence of any periodic component to these simulated rates.

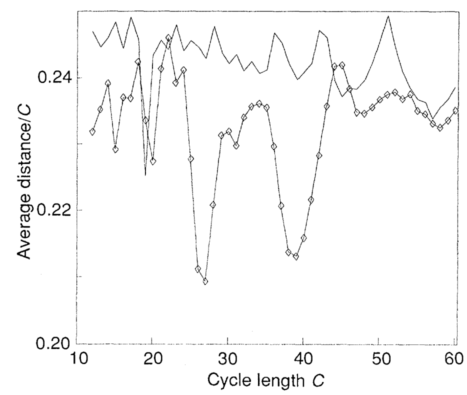

Fig. 7. The average (line marked with diamonds) of the 40 distances from the 40 Harland time points from 253 to 5.1 m.y. included in this study (*6*) to the nearest grid point on the best fitting cycle C (with phase chosen to minimize this average); the vertical scale gives these averages divided by the cycle length C. The other line shows the corresponding average distances (again relative to C) for the 40 equally spaced time points from 253 to 5.1 m.y.

$R_i > R_{i+1}$, (iii) iteratively computed $g(C)$ for each C, and (iv) then found D. Only in 17 of the 3000 series, or 0.6% of the cases, was D smaller than -1.05, giving an estimated P value of 0.0057, with an estimated standard error of 0.0014. We can thus be confident that $P < 0.01$ and that it is quite unlikely that a series of random rates would give a value for $D = \min d(C)$ as low as -1.05. Figure 2 shows an estimate of the probability density of D based on this simulation experiment. Thus far the analysis essentially duplicates that of Raup and Sepkoski (*2*), and it reaches the same conclusion: the series is not random.

This conclusion does not depend on the particular distribution used, namely, uniform from 0 to 1. Because the locations (and hence the distribution) of the peaks are unaffected by monotone transformations of the rates, exactly the same conclusion would be reached with any other distribution for the rates. Indeed, only the exchangeability of the rates is important, and thus the same conclusion would be reached if the original data values had been randomly rearranged 3000 times (*11*).

In the simulation study we kept track of both D and the corresponding cycle length C that gave this minimum value. Figure 3 shows a scatterplot of D versus C for the experiments; the lack of any pronounced relation confirms that subtracting $0.25C - 1.60$ from $g(C)$ does not favor any particular cycle length, although the lower dispersion for cycles less than 20 m.y. in duration effectively prevents their appearance as significantly low deviations. The surprising feature of these data only becomes apparent when the marginal distribution of C (Fig. 4) is examined: there is a strong tendency for random series to show their best fit for $C = 26$. Indeed, more than one-quarter (27%) of the best fitting cycles greater than 20 m.y. occurred at $C = 26$ or $C = 27$. That is, although the observed value of D is extremely low compared with this reference set, the observed best fitting period of 26 m.y. is that most probable for random series.

The test's intrinsic bias toward 26 m.y. With series using randomly generated rates at the predetermined t_i, the time scale exerts an intrinsic bias toward a best fitting period of 26 m.y., although the calibration of the test of course, is such that these deviations tend not to be significant. This suggests, however, that the effect may persist for nonrandom series and raises the possibility that for other models the test may be biased, that is, it may tend to indicate significant deviations at 26 m.y. even when there is no periodic component of 26 m.y. We performed a series of simulation experiments and found this to be the case. Specifically, if the rates exhibit a particular type of positive serial correlation, they will frequently produce significant deviations, and the most probable best fitting cycle length is 26 m.y. Furthermore, if the rates are periodic, they will tend to indicate a cycle of 26 m.y. regardless of the actual period unless the periodic signal is strong relative to the noise. Thus, while the test is valid as a means of testing the hypothesis that the rates are totally random, it cannot be used to discriminate among alternative models; it cannot even be used as a way to estimate the period length among periodic models.

We investigated three classes of models. If we let $\{U_i\}$ $(i \geq 0)$ represent a set of independent random variables, uniformly distributed from 0 to 1, then the models for the rates $\{R_i\}$ and the cases investigated are as follows.

1) Autoregressive models, θ = −0.1, 0.1, 0.2, 0.3, 0.4:

$$R_1 = U_1$$
$$R_i = \theta(R_{i-1} - 0.5) + U_i,\ i \geq 2.$$

2) Moving average models, θ = 0.25, 0.5, 1.0:

$$R_i = \theta U_{i-1} + U_i,\ i \geq 1.$$

3) Periodic models, periods P = 26, 30; θ = 0.25, 0.5, 1.0, 2.5:

$$R_i = (2\theta/5)\sin(2\pi t_i/P) + U_i,\ i \geq 1.$$

All three models reduce to random series if θ = 0. In all cases, 1000 series of length 40 were generated.

All models are intended to represent at least qualitatively plausible scenarios for series of extinction rates. Autoregressive models may be thought of as capturing lasting, but diminishing, environmental effects (*12*). Moving averages result from measurement error, as when extinctions are misallocated or fractionally allocated to two adjacent stages. Finally, periodic models represent potential astronomical causes (Nemesis "death stars," "Planet X," and so forth).

The significance of the observed deviation −1.05 was evaluated in all cases; the estimated P values are given in Table 1. The data in Table 1 suggest that, not only is the observed value $D = -1.05$ inconsistent with the model of a random series (θ = 0.0), it is also implausible for autoregressive models for the range of θ considered. It is, however, consistent with either moving average or periodic models if θ is sufficiently large.

As an aid to interpreting these results for periodic models, we note that for $\theta = 1.0$ the standard deviations of the purely periodic component and of the noise component are approximately equal: the signal-to-noise ratio is about 1.0 (*13*). For data such as the extinction data, the noise component may be expected to be large; indeed, comparison of the data series (*1*) with simulated series suggests that, even if there is a periodic component, a signal-to-noise standard deviation ratio on the order of 1:4 or 1:2 is plausible. For the models considered, this translates into a range of θ from 0.25 to 0.5. For such θ the test is not powerful at all; it is unlikely to detect the periodic signal.

Even more striking is the observation that, for signal-to-noise ratios in the range 1:4 to 1:2, the test is unlikely to indicate the correct period as that of best fit. The periodic signals of this strength are simply not sufficiently strong to override the bias toward 26 m.y. that is intrinsic in the time scale. Figure 5 shows the distribution of best fitting cycle lengths for 1000 series simulated from a model with a period of 30 m.y. and a signal-to-noise ratio of 1:4. Essentially the same results were obtained by repeating the experiment with a 15-m.y. phase shift. The bias toward a 26-m.y. period is pronounced.

The sensitivity of the test to moving average models is particularly intriguing, since these models may be considered plausible ones for the series of extinction rates because they may be viewed as modeling a type of measurement error prevalent in these data. For example, if actual extinctions occur totally at random but 50% of each epoch's actual extinctions are mistakenly recorded as belonging to the previous epoch (due to the lack of preservation of specimens), then the recorded series will be a moving average with $\theta = 1.0$. Figure 6 shows the distribution of C for the simulated moving average series with $\theta = 1.0$. If only 33% of the extinctions are misrecorded, $\theta = 0.5$. This is a simple model that qualitatively reflects what has been called the Signor-Lipps effect (*14*). The bias intrinsic to the time scale also exerts itself here and tends to favor 26 m.y. as the best fitting cycle.

The role of the time scale. We repeated part of this study for several other time scales, including the Odin scale (*7*), the Snelling scale (*8*), and an equally spaced (by 6.356 m.y.) grid of 40 time points from 5.1 to 253 m.y. The results for the Odin and Snelling scales differ little from those reported for the Harland scale; the preference for the exact value of 26 m.y. is less marked than for the Harland scale, but in both cases the propensity for random series to produce best fitting cycles in the neighborhood of 26 m.y. remains, as does the sensitivity to moving average alternatives. The situation for the equally spaced grid is markedly different for random series, however. For random series with equally spaced time points, the preference for 26 m.y. disappears, and most best fitting cycles are in the range 12 to 18 m.y., a result quite at variance with that for the Harland scale (*15, 16*). Since most of the Harland intervals are nearly equal (30 of the 40 are between 4 and 7 m.y.), this suggests that the sharp peak in Fig. 4 is due to the particular spacing of the small number of stratigraphic stages of long duration, especially the spacing of the Albian (113 to 97.5 m.y.; duration, 15.5 m.y.), the Campanian (83 to 73 m.y.; duration, 10 m.y.), the middle Eocene (50.5 to 42.0 m.y.; duration, 8.5 m.y.), and the early Miocene (24.6 to 14.4 m.y.; duration, 10.2 m.y.). Figure 7 shows how this unequal spacing favors certain periods: it shows the averages of the 40 distances from the time points t_i to the nearest cyclic grid mark $\hat{P}_i$ for the best fitting cycles from 12 to 60 m.y. for both the Harland and equally spaced scales. The best fitting cycles of lengths 26 and 27 m.y. are about 15% closer to the Harland time points on average than they are to the corresponding points on the equally spaced grid. This sensitivity to these particular stages has been verified by further simulation: even if only the Albian is subdivided, the propensity for cycles of 26 m.y. is considerably diminished. We thus conclude that the sharp propensity for a cycle of 26 m.y. for random series is an artifact of the particular spacing of a small number of long stages of the Harland scale.

The other results carry over more directly to alternative scales. The test's sensitivity to moving average alternatives is as great for the Odin and Snelling scales as for the Harland scale. More surprising, this sensitivity to moving average alternatives reintroduces a pseudoperiod in the neighborhood of 26 m.y. for the equally spaced grid scale. Despite the change in results for random series when the time scale is made more uniform, the behavior for moving average models is not greatly changed if the equally spaced grid scale is substituted for the Harland scale with a moving average process with $\theta = 1.0$. Figure 6 is nearly identical to the comparable figure for a simulation of a moving average process based on the equally spaced grid scale.

The nature of the test's sensitivity to the time scale and moving average models has several implications as to what should be expected if the data or the time scale (or both) are refined. First, improvements in the database may be expected to have little effect on the conclusions as long as the time scale is unchanged, unless the stochastic nature of the series is affected. This prediction is born out by the results of Raup and Sepkoski (*2*). They found little change in the results with improved family data, whereas genera data (which would be more susceptible to lack of preservation of specimens and hence have a stronger "moving average" aspect) gave stronger evidence of a cycle. Also, if the time scale is refined by subdividing a few large intervals, no major changes in the results are expected if there is a moving average characteristic to the data, as would be anticipated here. This prediction is supported by the results of Sepkoski (*17*). Indeed, if the subdivision is not accompanied by a refinement of the data, then no change for random series would be expected since the peaks of any series are undisturbed by linearly interpolating values between those observed.

We conclude the following. (i) The nonparametric test, properly calibrated, is a valid test of the null hypothesis of a random series. (ii) The test is not very powerful, however, and can only be counted on to detect strong periodic signals (signal-to-noise standard deviation ratio $\approx$2:1). (iii) The test is sensitive to nonperiodic alternatives as well, in particular to moving average processes. Since moving average processes could plausibly arise from measurement problems of the type known to occur with extinction data, this should be considered a likely explanation for the significant fit to an apparent cycle of 26 m.y. (iv) Even if the signal is periodic, biases intrinsic to the time scale render the test totally unreliable as a means of estimating the period, unless the signal is quite strong.

REFERENCES AND NOTES

1. D. M. Raup and J. J. Sepkoski, Jr., *Proc. Natl. Acad. Sci. U.S.A.* **81**, 801 (1984).
2. ———, *Science* **231**, 833 (1986).
3. D. M. Raup, *Nature (London)* **314**, 341 (1985).
4. R. B. Stothers, *ibid.* **322**, 444 (1986).
5. ———, *Astron. Astrophys.* 77, 121 (1979).
6. W. B. Harland *et al.*, *A Geologic Time Scale* (Cambridge Earth Sciences Series, Cambridge Univ. Press, Cambridge, England, 1982).
7. G. S. Odin, *Numerical Dating in Stratigraphy* (Wiley, New York, 1982).
8. N. J. Snelling, Ed., *The Chronology of the Geological Record* (Blackwell, Oxford, 1985).
9. J. J. Sepkoski, Jr., and D. M. Raup, in *Dynamics of Extinction*, D. K. Elliott, Ed. (Wiley, New York, 1986), chap. 1.
10. Another possible approach would be to attempt to estimate the actual 1% joint lower percent points of $g(C)$, which are not exactly linear in C, and subtract them from $g(C)$. This was not done both because of the great difficulty of finding them [due to the correlation of simulated $g(C)$ for different C, they are not simply related to the pointwise 1% points] and because this would not reflect the actual way in which the test has been performed in other studies [for example, the solid 99% lines of figure 2 of (*2*) correspond to straight lines on our scale].
11. In (*1*), Raup and Sepkoski report that they performed two kinds of simulations, those that randomly permuted the rates and those that randomly

permuted the rates but rejected the permutation if the number of peaks did not agree with that for their data. They found that this conditioning did not affect the results, as would be expected on the basis of the relative insensitivity of the number of peaks to the stochastic structure of the series [see S. M. Stigler, *Am. Stat.* **40**, 111 (1986)]. Indeed, insofar as there is any relation between the distribution of the number of peaks and the stochastic or periodic structure of the series, it is inappropriate to condition on the number of peaks if the goal is to learn about the relative likelihood of different alternative models. In some of the other studies reported (*9*), the series of peaks has been culled to limit attention to "significant" peaks before applying the randomization test. If this culling were based only on information in adjacent stages it would be essentially equivalent to conditioning on a smaller number of peaks; a valid test could be based on such a culled series, although a study of its sensitivity to alternative models would be difficult. Nevertheless, when the culling looks beyond adjacent stages to nearest minima [as was done in (*9*)], the simple randomization test will be biased toward periodic alternatives.

12. One extreme form of an autoregressive process was suggested in an earlier critique of this test by A. Hoffman [*Nature (London)* **315**, 659 (1985)] and by A. Hoffman and J. Ghiold [*Geol. Mag.* **122**, 1 (1985)]. In particular, Hoffman and Ghiold considered a random walk model, which is an autoregressive process with $\theta = 1.0$. Such a model implies a long system memory and has not been widely accepted.
13. Because the variance of the sine of a random angle is six times that of U_i, the ratio of the standard deviation of the periodic signal to that of the random noise is 0.98θ, and hence θ may be conveniently thought of as the signal-to-noise ratio here, where we use this term to refer to the ratio of the respective standard deviations.
14. D. M. Raup, *Science* **231**, 1528 (1986); based on P. W. Signor III and J. H. Lipps [in *Geological Implications of Impacts of Large Asteroids and Comets on the Earth, Conference on Large Body Impacts and Terrestrial Evolution*, L. T. Silver and P. H. Schultz, Eds. (Geological Society of America Special Papers, no. 190, Boulder, CO, 1982), pp. 291–296]. The manner in which misrecorded extinctions generate a moving average process of the type we consider is most easily seen where the stratigraphic stages are of equal duration. In that case, if in each stage a fraction $\theta/(1 + \theta)$ of that stage's extinctions are misrecorded as occurring in the previous stage and the remaining fraction $1/(1 + \theta)$ are correctly recorded, then the series of recorded extinctions (and hence the rates) will be proportional to a "moving average" of the type we consider, with weight θ on the previous stage and weight 1 on the current stage. The same will be true for series of unequal stages if the fraction misrecorded is rescaled by the ratio of the successive stage lengths.
15. Thus the effect found here is not related to the "wrapping effect" noticed by T. M. Lutz [*Nature (London)* **317**, 404 (1985)] in a critique of the application of this test to magnetic reversal data. That effect, while real, is quite small here, as indeed it was in the magnetic reversal case. D. M. Raup (*ibid.*, p. 384) has noted that the extinction data do not exhibit the type of long-term variation that was responsible for the major effect at the basis of the convincing criticism Lutz made of the application of this test to the reversal data.
16. The sensitivity of the nonparametric test to patterns in the duration of stages helps explain why it tends to support different conclusions from those reached by analyses like that of J. A. Kitchell and D. Pena [*Science* **226**, 689 (1984)], which treat the observed rates as if they are equally spaced.
17. J. J. Sepkoski, Jr., *Lect. Notes Earth Sci.* **8**, 47 (1986).
18. Supported in part by NSF grant DMS-8601732. This manuscript was prepared with the use of computer facilities supported in part by NSF grants DMS-8601732 and DMS-8404941 to the Department of Statistics at the University of Chicago. We thank D. M. Raup, J. J. Sepkoski, Jr., M. Stein, and W. H. Wong for helpful discussions and criticism.

16 April 1987; accepted 6 August 1987

Atmospheric Trends in Methylchloroform and the Global Average for the Hydroxyl Radical

R. PRINN, D. CUNNOLD, R. RASMUSSEN, P. SIMMONDS, F. ALYEA, A. CRAWFORD, P. FRASER, R. ROSEN

Frequent atmospheric measurements of the anthropogenic compound methylchloroform that were made between 1978 and 1985 indicate that this species is continuing to increase significantly around the world. Reaction with the major atmospheric oxidant, the hydroxyl radical (OH), is the principal sink for this species. The observed mean trends for methylchloroform are 4.8, 5.4, 6.4, and 6.9 percent per year at Aldrigole (Ireland) and Cape Meares (Oregon), Ragged Point (Barbados), Point Matatula (American Samoa), and Cape Grim (Tasmania), respectively, from July 1978 to June 1985. These measured trends, combined with knowledge of industrial emissions, were used in an optimal estimation inversion scheme to deduce a globally averaged methylchloroform atmospheric lifetime of 6.3 (+1.2, −0.9) years (1σ uncertainty) and a globally averaged tropospheric hydroxyl radical concentration of $(7.7 \pm 1.4) \times 10^5$ radicals per cubic centimeter (1σ uncertainty). These 7 years of gas chromatographic measurements, which comprise about 60,000 individual calibrated real-time air analyses, provide the most accurate estimates yet of the trends and lifetime of methylchloroform and of the global average for tropospheric hydroxyl radical levels. Accurate determination of hydroxyl radical levels is crucial to understanding global atmospheric chemical cycles and trends in the levels of trace gases such as methane.

METHYLCHLOROFORM IS A LONG-lived atmospheric species whose only known sources are anthropogenic. It is widely used in industry as a solvent for degreasing and in other applications; its global production and use accelerated significantly in the mid-1970s as it replaced the increasingly regulated solvent trichloroethylene (*1*). Its concentration has rapidly increased worldwide (*1–3*). This increase is of considerable concern because methylchloroform is a significant stratospheric source of atomic chlorine and chlorine monoxide (*4*) that can catalytically destroy stratospheric ozone (*5*), and because it is one of the greenhouse gases that may contribute to future climate change (*6*).

The principal recognized atmospheric sink for methylchloroform (CH_3CCl_3) is the reaction (*7, 8*)

$$CH_3CCl_3 + OH \rightarrow CH_2CCl_3 + H_2O \quad (1)$$

The global rate of loss of CH_3CCl_3, which can be deduced from its known industrial emissions and observed global trends, can be used to deduce accurately an appropriately weighted globally averaged tropospheric concentration for the hydroxyl radical (OH) (*1–3, 8*). Knowledge of precise OH concentrations is crucial because OH is widely recognized as the major gas-phase oxidant in clean tropospheric air. Thus it plays a central role in tropospheric chemistry, including chemical destruction of a wide range of species of anthropogenic and natural origin that are important radiatively or chemically or both (for example, CO, CH_4, NO_x, and SO_2) (*9*).

Since 1978 we have performed frequent (4 to 12 measurements per day) real-time gas chromatographic measurements of methylchloroform at stations throughout the world, first as part of the Atmospheric Lifetime Experiment (ALE) and (since late 1984) as part of the Global Atmospheric Gases Experiment (GAGE) (*1, 10*). The surface measurement stations are located at coastal sites remote from industrial and urban sources and are designed to measure accurately the tropospheric trends of trace gases whose lifetimes are long compared to global tropospheric mixing times (*11*). In this report we present and interpret ALE-GAGE data obtained over the 7-year period from July 1978 to June 1985 which, primarily because of the greater number of measurements, provide much more accurate determinations than previously possible (*1–3, 8*) of the global trend and atmospheric destruction time for methylchloroform and for the globally averaged OH concentration.

Monthly mean volume mixing ratios for methylchloroform were computed from the approximately 100 to 400 measurements made each month at each of the five ALE-

R. Prinn, Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139.
D. Cunnold and F. Alyea, School of Geophysical Sciences, Georgia Institute of Technology, Atlanta, GA 30332.
R. Rasmussen and A. Crawford, Institute of Atmospheric Sciences, Oregon Graduate Center, Beaverton, OR 97005.
P. Simmonds, Department of Geochemistry, University of Bristol, Bristol, BF8-1TS, United Kingdom.
P. Fraser, Division of Atmospheric Research, Commonwealth Scientific and Industrial Research Organization, Aspendale, 3195 Victoria, Australia.
R. Rosen, Atmospheric and Environmental Research, Inc., Cambridge, MA 02139.