

Confidence intervals for pulsed mass extinction events

Steve C. Wang and Philip J. Everson

Abstract.—Many authors have proposed scenarios for mass extinctions that consist of multiple pulses or stages, but little work has been done on accounting for the Signor-Lipps effect in such extinction scenarios. Here we introduce a method for computing confidence intervals for the time or stratigraphic distance separating two extinction pulses in a pulsed extinction event, taking into account the incompleteness of the fossil record. We base our method on a flexible likelihood ratio test framework that is able to test whether the fossil record is consistent with any extinction scenario, whether simultaneous, pulsed, or otherwise. As an illustration, we apply our method to a data set on marine invertebrates from the Permo-Triassic boundary of Meishan, China. Using this data set, we show that the fossil record of ostracodes and that of brachiopods are each consistent with simultaneous extinction, and that these two extinction pulses are separated by 720,000 to 1.2 million years with 95% confidence. With appropriate data, our method could also be applied in other situations, such as tests of origination patterns, coordinated stasis, and recovery after a mass extinction.

Steve C. Wang and Philip J. Everson. Department of Mathematics and Statistics, Swarthmore College, Swarthmore, Pennsylvania 19081. E-mail: scwang@swarthmore.edu and E-mail: everson@swarthmore.edu

Accepted: 30 January 2007

Introduction

Most previous work on the Signor-Lipps effect (Signor and Lipps 1982) has focused on estimating the time or stratigraphic position of a mass extinction event in which taxa are thought to have gone extinct simultaneously (Strauss and Sadler 1989; Springer 1990; Marshall 1995; Marshall and Ward 1996; Solow 1996; Wang and Marshall 2004). Many authors, however, have proposed mass extinction scenarios in which taxa go extinct in distinct pulses or stages (e.g., Stanley and Yang 1994; Knoll et al. 1996; Paul et al. 1999; McGhee 2001; Isozaki 2002; Keller et al. 2003; Xie et al. 2005). Little methodology has been developed for quantifying characteristics of such pulsed extinction events. In this paper, we introduce a method for estimating the separation or duration of a two-pulse extinction event. Using a likelihood-based method, we demonstrate how to compute a confidence interval for the elapsed time or stratigraphic distance separating the two extinction pulses.

To derive our confidence interval, we take advantage of the fact that confidence intervals are intrinsically related to hypothesis tests. Therefore, we first describe a flexible likeli-

hood-based framework for testing any mass extinction scenario, whether simultaneous, pulsed, or otherwise. Using this hypothesis test, we generate a joint confidence region that shows the times or locations of the two extinction pulses that are consistent with the fossil record. We then use this region to graphically determine a confidence interval for the time or stratigraphic distance separating the two pulses. We also derive a simple algebraic formula for the confidence interval. To illustrate, we apply our method to a data set on marine invertebrates from the Permo-Triassic boundary of Meishan, China (Jin et al. 2000), and estimate that end-Permian ostracode and brachiopod extinctions were separated by 720,000 to 1.22 million years with 95% confidence.

Ostracodes and Brachiopods in the End-Permian Extinction

The end-Permian extinction was the most severe mass extinction of the Phanerozoic, with over 90% of all marine species being eliminated (Jin et al. 2000). The best-studied Permo-Triassic marine boundary sections in the world are those at Meishan in South China

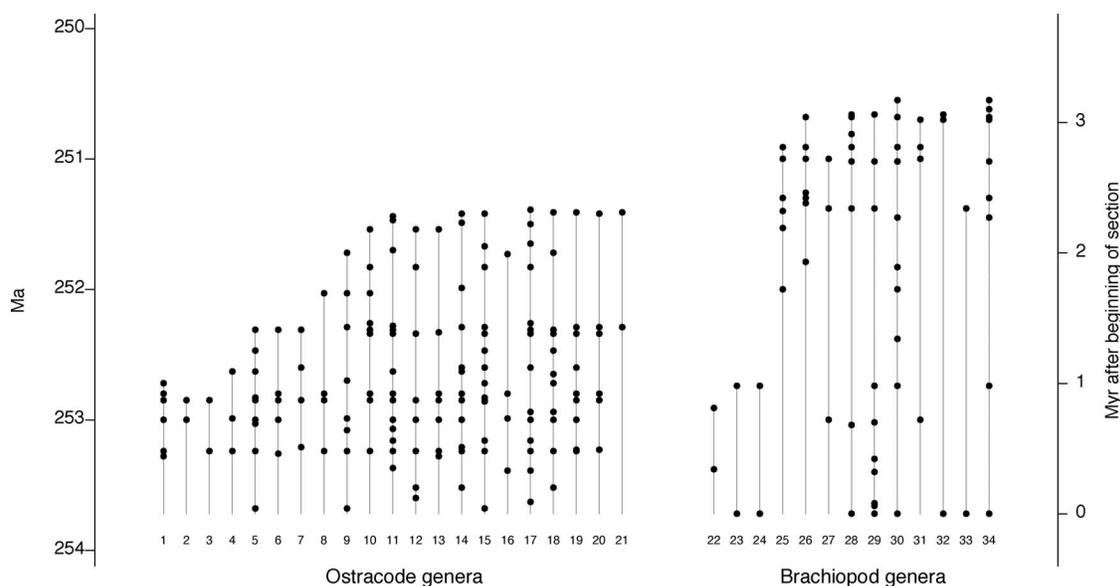


FIGURE 1. Stratigraphic ranges of ostracode and brachiopod genera used in this study, redrawn from Figure 2 of Jin et al. (2000). All 21 ostracode genera and 13 brachiopod genera given by Jin et al. (2000: Fig. 2) are included; no genera have been selectively excluded. Ostracodes are numbered as genera 1–21; brachiopods are numbered as genera 22–34. Left axis gives time in Ma as given by Jin et al. (2000); right axis gives time in Myr after the beginning of the section.

(Jin et al. 2000). To demonstrate our confidence interval method, we use data on stratigraphic ranges of ostracodes and brachiopods from Meishan (Jin et al. 2000; transcribed from their Figure 2). These data give the times of fossil finds of 21 genera of ostracodes and 13 genera of brachiopods, ranging from 250.55 Ma to 253.72 Ma (Fig. 1). In subsequent analyses, we refer to these genera by indices from 1 to 34, with genera 1–21 referring to the ostracodes and 22–34 to the brachiopods.

For our analyses, it is necessary to express the data in units of time after the beginning of the section (or equivalently, stratigraphic distance above the base of the section), rather than as absolute ages. We therefore subtract the time of each fossil find from 253.72 Ma, the time corresponding to the base of the section (Jin et al. 2000: Fig 2). For instance, a fossil brachiopod dated to 251.02 Ma would be recorded as $253.72 - 251.02 = 2.70$ Ma after the beginning of the section. Thus our time or distance values increase toward the top of the section.

From these data (Fig. 1), several questions naturally arise:

1. Did ostracodes go extinct simultaneously, and if so, when? Did brachiopods go extinct simultaneously, and if so, when?
2. If both ostracodes and brachiopods went extinct simultaneously, did they do so at the same time? If so, when?
3. If not, how much time separated their extinctions?

We can use existing methods (Solow 1996; Wang and Marshall 2004) to answer questions (1) and (2). In this paper, we develop a method for answering question (3), which has not previously been addressed. Our approach is to calculate a confidence interval for the time separating the extinction of ostracodes and the extinction of brachiopods. Our confidence interval method is based on a flexible hypothesis test framework that can be adapted to model a variety of extinction scenarios. We therefore begin by describing this hypothesis test.

A Flexible Framework for Specifying Extinction Hypotheses

Here we describe the hypothesis test on which our confidence interval method is

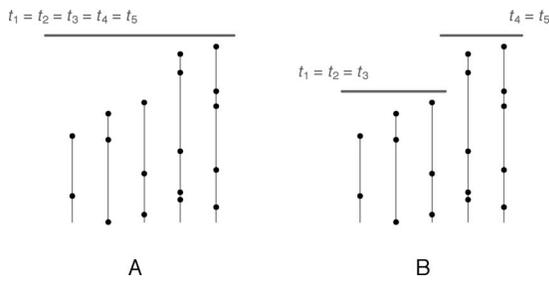


FIGURE 2. Examples of extinction scenarios that can be tested using our framework. A, Simultaneous extinction of all five taxa at time t_1 . B, Two-pulse (stepwise) extinction; taxa 1–3 go extinct at time t_1 and taxa 4 and 5 at time t_4 .

based. Most previous work has focused on testing whether or not extinctions were simultaneous (Strauss and Sadler 1989; Springer 1990; Marshall 1995; Marshall and Ward 1996; Solow 1996; Wagner 2000; but see Solow and Smith 2000). Simultaneous extinction is taken as the null hypothesis because it is mathematically tractable, and because it offers a single explanation for multiple extinctions and is therefore parsimonious.

In order to develop our confidence interval methodology, we require a hypothesis test framework that can test pulsed extinction hypotheses as well as simultaneous extinction hypotheses. Here we introduce such a framework that is able to test any kind of extinction scenario, whether simultaneous, pulsed, or gradual, or some combination thereof. We later describe how to use the hypothesis test to calculate confidence intervals for the time separating the extinction pulses.

We first define the notation we will use. Let G denote the number of genera in a section that are victims of the mass extinction; here $G = 34$ ($= 21$ ostracode genera + 13 brachiopod genera). Let $\theta_1, \dots, \theta_G$ denote the unknown true extinction times of genera 1 through G , and let t_1, \dots, t_G denote the hypothesized extinction times we wish to test. (In our examples we refer to genera, but the method applies equally well to data at any taxonomic level. The method can also be used for data recorded as stratigraphic positions, as is usually the case, rather than time.)

Our goal is to test whether the extinctions in fact occurred at the hypothesized extinction

times t_1, \dots, t_G . Our null hypotheses thus have the form

$$H_0: \theta_1 = t_1, \theta_2 = t_2, \dots, \theta_T = t_G,$$

which we test against the alternative hypothesis that t_1, \dots, t_G do not represent the true extinction times. Our model can accommodate a variety of extinction scenarios, depending on the values of the t 's that are specified.

Simultaneous Extinction.—If all the t 's are equal ($t_1 = t_2 = \dots = t_G$), then H_0 hypothesizes that a sudden extinction occurred, with all taxa going extinct simultaneously at the common time t_1 (Fig. 2A). Such a scenario might be expected in the case of a single catastrophic cause, such as a bolide impact (e.g., Alvarez et al. 1980).

Pulsed (Stepwise) Extinction.—Suppose we want to test whether a two-pulse extinction occurred, with the first k taxa going extinct at time t_1 and the remaining $(G - k)$ taxa at time t_{k+1} . This can be modeled by a null hypothesis in which $t_1 = t_2 = \dots = t_k$ and $t_{k+1} = t_{k+2} = \dots = t_G$, with $t_k \neq t_{k+1}$. More generally, any number of extinction pulses may be specified by choosing the t 's appropriately. Figure 2B shows an example in which three taxa are hypothesized to go extinct at time $t_1 = t_2 = t_3$ and two others at time $t_4 = t_5$. Such a scenario might be expected in the case of an extinction caused by multiple catastrophic events (e.g., Stanley and Yang 1994; Knoll et al. 1996; Paul et al. 1999; McGhee 2001; Isozaki 2002; Keller et al. 2003; Xie et al. 2005).

Other Extinction Scenarios.—A gradual extinction can be modeled in our framework by using unequal t 's ($t_1 \neq t_2 \neq \dots \neq t_G$). Combinations of simultaneous, pulsed, and gradual extinctions can be accommodated as well. In practice, however, it is difficult to test such scenarios because t_1, \dots, t_G must be specified in advance, which is rarely possible for scenarios involving gradual extinctions. We therefore do not discuss such scenarios further.

In general, hypothesis tests require that the null hypothesis be specified in advance. In our case, this means that the hypothesized extinction times t_1, \dots, t_G must be specified before observing the data. Such hypothesized extinction times may be known if there is independent geophysical or geochemical evidence, or

if times are hypothesized from one data set and tested using a different data set. Even when hypothesized times are not known in advance, however, we show below that simultaneous extinction scenarios can still be tested in our framework. Furthermore, as we also show, the hypothesis test for pulsed extinctions can be used to compute confidence intervals for the time separating pulses even when extinction times are not known in advance.

Calculating the Likelihood Ratio

In this section we show how to use the likelihood ratio (Casella and Berger 2002) to test hypotheses specified using the framework described above.

Likelihood-based models have been used in the paleontological literature in many contexts (e.g., Solow 1996; Huelsenbeck and Rannala 1997; Foote 1997; Solow and Smith 1997, 2000; Wagner 2000; McConway and Sims 2004). Solow (1996) used a likelihood-based approach to testing for simultaneous extinctions. Our approach is similar to Solow's, but we extend the methodology to pulsed extinctions. Wagner (2000) described a likelihood-based approach to testing the number of pulses in a mass extinction event. Our approach differs from Wagner's in several respects: (1) Wagner does not extend his method to calculating confidence intervals, as we do here; (2) Wagner's model applies to data collected by discrete sampling, whereas ours applies to continuous sampling; (3) Wagner's approach focuses on testing the number of biological or sampling parameters, whereas ours focuses on how much the data deviate from the expectations of particular hypotheses.

For simplicity, we assume continuous sampling and uniform preservation and recovery of fossil taxa. Although these are strong assumptions, they are reasonable for some data sets (Macellari 1986) and have commonly been assumed in the literature (Strauss and Sadler 1989; Springer 1990; Marshall 1995; Marshall and Ward 1996; Solow 1996; Solow and Smith 2000). For the Meishan ostracode and brachiopod data, Jin et al. (2000) state that collection intensity was uniform with no sampling gaps.

Recovery potential may nonetheless be non-uniform owing to changes in abundance, but such variation is difficult to quantify.

Notation.—Let n_i denote the number of fossil finds or horizons for genus i ; for example, in Figure 1 $n_1 = 6$, $n_2 = 2$, $n_3 = 2$, etc. Let x_{ij} denote the time of the j th fossil find for genus i ; the index i runs from genera 1 through G , and the index j runs from fossil finds 1 through n_i . For example, for genus 1 in Figure 1, we have $x_{11} = 1.00$, $x_{12} = 0.92$, $x_{13} = 0.87$, etc. (the units are millions of years (Myr) after the beginning of the section). For notational convenience, we will sometimes use x_i to denote the vector of fossil finds for genus i . That is, $x_i = \{x_{i1}, x_{i2}, \dots, x_{in_i}\}$; the boldface indicates that x_i is a vector. Thus x_1, \dots, x_G denote vectors giving the times of fossil finds for genera 1 through G . Also let y_i denote the time of the latest (or highest) fossil find for genus i ; that is, $y_i = \max_j(x_{ij})$ for genus i . As defined earlier, θ_i denotes the true extinction time of genus i .

The Likelihood Function.—The likelihood function, denoted $L(t_1, \dots, t_G | x_1, \dots, x_G)$, gives the likelihood of the hypothesized extinction times t_1, \dots, t_G , given the observed fossil finds x_1, \dots, x_G . By definition, $L(t_1, \dots, t_G | x_1, \dots, x_G)$ is numerically equal to the probability (to be precise, the joint probability density function) of observing the fossil finds x_1, \dots, x_G if t_1, \dots, t_G were the true extinction times. However, the likelihood differs conceptually from the joint probability in that it is considered a function of the hypothesized times (conditional on the data), whereas the joint probability is considered a function of the data (conditional on the hypothesized times). That is, the likelihood measures how well a particular set of hypothesized extinction times t_1, \dots, t_G (as opposed to some other set of extinction times) accounts for the fossil finds x_1, \dots, x_G . On the other hand, the joint probability measures the chances of observing a particular set of fossil finds x_1, \dots, x_G (as opposed to some other set of fossil finds) assuming the extinction truly occurred at times t_1, \dots, t_G .

The Likelihood Ratio Statistic.—To measure the strength of evidence against the null hypothesis, we use the likelihood ratio statistic, denoted by λ :

$$\lambda = \frac{\text{Value of } L(t_1 \dots t_G | \mathbf{x}_1 \dots \mathbf{x}_G) \text{ using the hypothesized times } t_1 \dots t_G \text{ specified by } H_0}{\text{Maximum value } L(t_1 \dots t_G | \mathbf{x}_1 \dots \mathbf{x}_G) \text{ can attain using any possible values of } t_1 \dots t_G}.$$

The likelihood ratio statistic takes on values between 0 and 1 and measures how likely the hypothesized values of t_1, \dots, t_G are, compared to the most likely possible values of t_1, \dots, t_G . If the hypothesized values of t_1, \dots, t_G are almost as likely as any values of t_1, \dots, t_G could possibly be, then these hypothesized values explain the observed data well, and we should not reject the null hypothesis. In this case, λ will be large (close to 1). On the other hand, if the hypothesized values of t_1, \dots, t_G are not likely, in the sense that they do not explain the observed data adequately, then we should reject the null hypothesis. In this case, λ will be small (close to 0). Therefore, we calculate λ for the hypothesized values of t_1, \dots, t_G , and we reject the null hypothesis if λ is sufficiently small. What constitutes “sufficiently small” depends on the alpha level (the probability of a Type I error or false positive).

From the assumption of uniform preservation and recovery, it follows that for each genus i , the times of the n_i fossil finds X_{ij} are distributed uniformly over the interval $[0, \theta_i]$. Using this result, we can show (see Appendix 1) that the likelihood ratio statistic is equal to

$$\lambda = \left(\frac{y_1}{t_1}\right)^{n_1} \left(\frac{y_2}{t_2}\right)^{n_2} \dots \left(\frac{y_G}{t_G}\right)^{n_G} \quad \text{if } t_i \geq y_i \text{ for all genera } i. \quad (1)$$

Note that if $t_i \geq y_i$ is not satisfied, then $\lambda = 0$. This occurs if the hypothesized extinctions predate the latest fossil find, which is impossible and thus immediately disproves the null hypothesis.

The Sampling Distribution of the Test Statistic.—We reject H_0 if the value of λ is sufficiently small. To ascertain which values of λ are “sufficiently small,” we must determine the sampling distribution of λ when the null hypothesis is true. Rather than determine the sampling distribution of λ , however, it is mathematically simpler to determine the sampling distribution of $-2 \log \lambda$. Because $-2 \log \lambda$ is a strictly monotonic function of λ , a hypothesis test based on $-2 \log \lambda$ will be

equivalent to one based on λ . Note, however, that small values of λ correspond to large values of $-2 \log \lambda$, so we reject H_0 if $-2 \log \lambda$ is sufficiently large—e.g., corresponding to a p -value less than 0.05.

Equation (1) can be written in the following equivalent form:

$$-2 \log \lambda = -2 \sum_{i=1}^G n_i \log \left(\frac{y_i}{t_i} \right) \quad \text{if } t_i \geq y_i \text{ for all genera } i. \quad (2)$$

It can be shown (see Appendix 2) that when H_0 is true, the test statistic $-2 \log \lambda$ has a chi-square distribution with $2G$ degrees of freedom. We therefore reject H_0 when

$$-2 \log \lambda > \chi^2_{2G,95}, \quad (3)$$

where $\chi^2_{2G,95}$ denotes the 95th percentile of the chi-square distribution with $2G$ degrees of freedom. Such percentiles of the chi-square distribution are commonly available in books and software packages.

An Illustration: Meishan Ostracodes and Brachiopods

To illustrate the likelihood ratio test and our confidence interval method, we apply our method to the data set of 21 ostracode and 13 brachiopod genera described above. Our goal is to compute a confidence interval for the time separating the ostracode and brachiopod extinction pulses. In order for this quantity to be meaningful, it is first necessary to establish that (1) all ostracode extinctions were simultaneous, (2) all brachiopod extinctions were simultaneous, and (3) the ostracode extinction pulse did not occur at the same time as the brachiopod extinction pulse. We therefore begin by using the likelihood ratio test to demonstrate that the fossil record at Meishan is consistent with a simultaneous extinction of all 21 ostracode genera and with a simultaneous extinction of all 13 brachiopod genera. Next, using the likelihood ratio test we show that the ostracode extinction pulse and the brachiopod extinction pulse do not appear to

be synchronous. Finally, we estimate the time separating the ostracode and brachiopod extinction pulses using our confidence interval method.

Hypothesis Test I: Ostracodes.—We first test the hypothesis that all 21 ostracode genera went extinct simultaneously. The corresponding null hypothesis in our framework is

$$H_0: \theta_1 = \theta_2 = \dots = \theta_{21} = t_{os} \quad (4)$$

where t_{os} denotes the common extinction time for the ostracode genera (i.e., $t_{os} = t_1 = t_2 = \dots = t_{21}$). However, what we want to test is not whether all genera went extinct simultaneously at some *particular* time t_{os} , but whether they went extinct simultaneously at *any* time—a hypothesis of the form

$$H_0: \theta_1 = \theta_2 = \dots = \theta_G.$$

In our framework, this hypothesis can be tested by setting t_{os} in equation (4) equal to the latest fossil find among the genera in question. This procedure, equivalent to the generalized likelihood ratio test (Casella and Berger 2002), maximizes the likelihood function over all possible common extinction times. For the Meishan data, the latest ostracode occurs at 2.33 Myr after the beginning of the section (an absolute date of 251.39 Ma), so we set $t_{os} = 2.33$ Myr in equation (4). Of all possible common extinction times, this value is the most consistent with the observed fossil record and therefore the most difficult to reject. Therefore, if our test rejects this value for a common extinction time, it would reject any other value of t_{os} as well. To be specific, any value of $t_{os} > 2.33$ would yield a larger value of $-2 \log \lambda$ and thus be rejected, and any value of $t_{os} < 2.33$ would make the observed data impossible and thus be rejected as well. We are thereby able to test a hypothesis of simultaneous extinction (occurring at any time) in our framework. This procedure is a continuous analogue of the discrete test given by Wagner (2000), and it is equivalent to the continuous test described by Solow (1996), although Solow does not generalize his result to pulsed extinctions as we do below.

To test the hypothesis given by equation (4), we calculate the test statistic $-2 \log \lambda$. As described above, this test statistic has a chi-

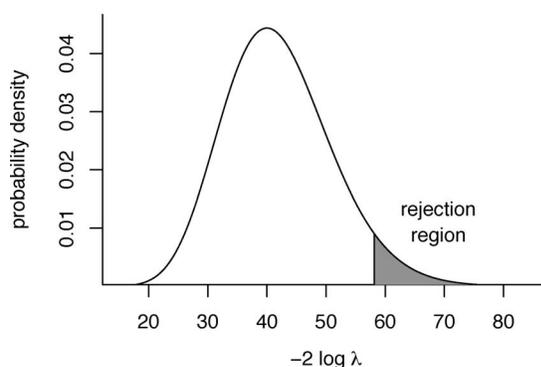


FIGURE 3. Sampling distribution of the test statistic $-2 \log \lambda$ under the null hypothesis for Meishan ostracodes. The sampling distribution is a chi-square distribution with $2G$ degrees of freedom, where $G = 21$ is the number of ostracode genera. Using a significance level of $\alpha = 0.05$, the rejection region (for which we reject the null hypothesis) comprises the highest 5% of values. This region corresponds to $-2 \log \lambda > 58.1$ and is shaded in gray.

square distribution with 42 ($= 2 \times 21$) degrees of freedom when the null hypothesis is true; this sampling distribution is shown in Figure 3. The 95th percentile of this distribution is $\chi^2_{42,95} = 58.1$. Therefore the rejection region for the hypothesis test (comprising the highest 5% of the area under the curve) occurs when $-2 \log \lambda \geq 58.1$ (shaded region in Fig. 3). For the ostracode data, the observed value of $-2 \log \lambda$ is 52.1, calculated using equation (2):

$$\begin{aligned} -2 \log \lambda &= -2 \sum_{i=1}^G n_i \log \left(\frac{y_i}{t_i} \right) \\ &\quad \text{if } t_i \geq y_i \text{ for all genera } i. \\ &= -2 \left[6 \log \left(\frac{1.00}{2.33} \right) + 2 \log \left(\frac{0.87}{2.33} \right) \right. \\ &\quad \left. + \dots + 2 \log \left(\frac{2.31}{2.33} \right) \right] \\ &= 52.1. \end{aligned}$$

As can be seen in Figure 3, this value is somewhat large but not large enough to fall in the rejection region (the corresponding p -value is 0.14). We therefore fail to reject H_0 and conclude that the fossil record of these 21 ostracode genera is consistent with a simultaneous extinction.

Hypothesis Test II: Brachiopods.—We next test the null hypothesis that all 13 brachiopod gen-

era went extinct simultaneously. The corresponding null hypothesis is

$$H_0: \theta_{22} = \theta_{23} = \dots = \theta_{34} = t_{br}$$

where t_{br} denotes the common extinction time for the brachiopod genera (i.e., $t_{br} = t_{22} = t_{23} = \dots = t_{34}$). The latest brachiopod occurrence was dated at 3.17 Myr after the beginning of the section (an absolute date of 250.55 Ma). Thus, following the logic of hypothesis test I, we set $t_{br} = 3.17$ Myr. Here the test statistic $-2 \log \lambda$ has a chi-square distribution with 26 degrees of freedom when the null hypothesis is true. The 95th percentile of this distribution is $\chi^2_{26,95} = 38.9$. For the brachiopod data, the observed value of $-2 \log \lambda$ is 21.3, calculated using equation (2), far below the cutoff needed to reject the null hypothesis ($p = 0.73$). We therefore fail to reject H_0 and conclude that the fossil record of these 13 brachiopod genera is consistent with a simultaneous extinction.

Hypothesis Test III: Simultaneous Extinction.—Because hypothesis tests I and II concluded that the fossil record of each group is consistent with a simultaneous extinction, we now test whether the ostracode and brachiopod extinction pulses were synchronous. In our framework, the corresponding null hypothesis is

$$H_0: \theta_1 = \theta_2 = \dots = \theta_{34} = 3.17$$

because 3.17 Myr after the beginning of the section is the time of the latest fossil find among any of the 34 genera. Under this null hypothesis, the test statistic $-2 \log \lambda$ has a chi-square distribution with 68 degrees of freedom, and the 95th percentile of this distribution is $\chi^2_{68,95} = 88.3$. The observed value of $-2 \log \lambda$ is 168.4, calculated using equation (2), far beyond the cutoff needed to reject the null hypothesis ($p = 0.00000000017$). We therefore reject H_0 and conclude that ostracodes and brachiopods did not go extinct at the same time.

Estimating the Time Separating Extinction Pulses: A Graphical Procedure

Because we have rejected a simultaneous extinction time for ostracodes and brachiopods, it is natural to ask how much time separated their two extinction pulses (Fig. 4).

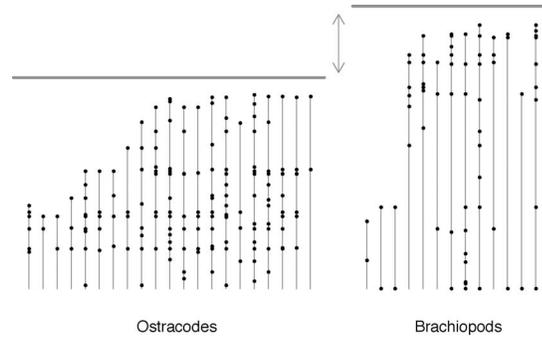


FIGURE 4. Stratigraphic ranges of ostracode and brachiopod genera. The fossil record is consistent with a simultaneous extinction of each group. However, we reject the hypothesis of synchronous extinction of ostracodes and brachiopods together. We therefore estimate the time separating the ostracode extinction pulse and the brachiopod extinction pulse (represented by arrow).

Here we introduce a method for creating a confidence interval to estimate this time, $(\theta_{br} - \theta_{os})$. We begin by defining a joint confidence region for the two extinction pulses, and we then use this region to graphically determine a confidence interval for the time separating the two extinction pulses. To do so, we exploit the relationship between hypothesis tests and confidence intervals: a confidence interval is the set of parameter values that would not be rejected by a hypothesis test.

Joint Confidence Region for θ_{os} and θ_{br} .—A joint confidence region is a two-dimensional analogue of a confidence interval, consisting of all points (t_{os}, t_{br}) for which the null hypothesis

$$H_0: \theta_1 = \theta_2 = \dots = \theta_{21} = t_{os}; \\ \theta_{22} = \theta_{23} = \dots = \theta_{34} = t_{br} \quad (5)$$

is not rejected at the corresponding alpha level. In other words, one can think of the joint confidence region as the set of all plausible values for the extinction times. To calculate the 95% joint confidence region, we test hypothesis (5) for all combinations of t_{os} and t_{br} , using an alpha level of $\alpha = 0.05$. A gray dot is plotted at the point (t_{os}, t_{br}) if the hypothesis test does not reject H_0 , indicating that for the values (t_{os}, t_{br}) hypothesis (5) is consistent with the observed fossil record. The 95% joint confidence region, consisting of all points (t_{os}, t_{br}) consistent with the observed fossil record, is shown as a gray triangular region in Figure 5.

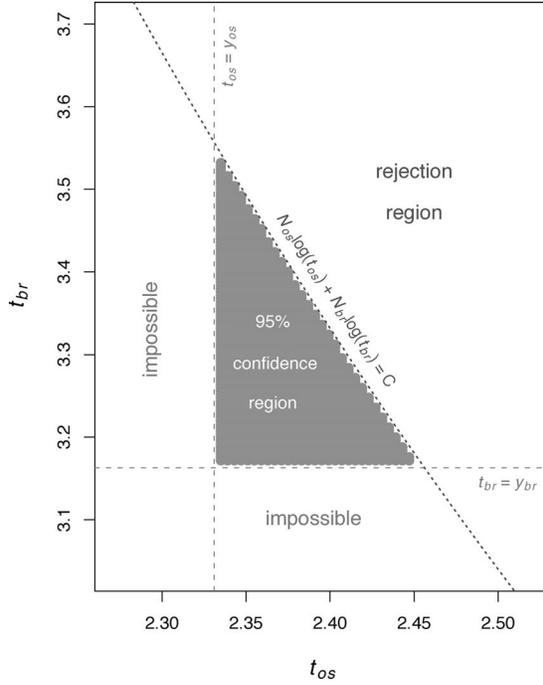


FIGURE 5. Joint (two-dimensional) 95% confidence region for θ_{os} and θ_{br} , the extinction times for Meishan ostracodes and brachiopods (data from Jin et al. 2000). Points in the 95% confidence region (gray shaded area) are values of (t_{os}, t_{br}) for which the fossil record is consistent with an ostracode extinction pulse at time t_{os} and a brachiopod extinction pulse at time t_{br} , using a hypothesis test with $\alpha = 0.05$. For instance, the point (2.4, 3.3) is shaded in gray, indicating that the fossil record is consistent with an ostracode extinction pulse at 2.4 Myr after the beginning of the section and a brachiopod extinction pulse at 3.3 Myr after the beginning of the section. The vertical dashed line marks the time (y_{os}) of the latest ostracode fossil find; the horizontal dashed line marks the time (y_{br}) of the latest brachiopod fossil find. The area to the left and below these dashed lines consists of impossible values for (t_{os}, t_{br}) because the genera were extant. The diagonal dotted curve (given by equation 7) is the boundary of the rejection region. The area above the dotted curve consists of values of (t_{os}, t_{br}) that are rejected by a hypothesis test with $\alpha = 0.05$ and are therefore implausible.

Determining the Confidence Interval for $(\theta_{br} - \theta_{os})$.—Figure 5 can be used to graphically determine a confidence interval for $(\theta_{br} - \theta_{os})$, the time separating the two extinction pulses. Note that t_{br} increases toward the top of Figure 5, and t_{os} decreases toward the left. Therefore, the difference $(t_{br} - t_{os})$ increases toward the top left of the graph. Hence the largest value of $(t_{br} - t_{os})$ that is consistent with the observed data is found at the top left vertex of

the joint confidence region. This point has coordinates $t_{os} = 2.33$ Myr and $t_{br} = 3.55$ Myr after the beginning of the section, corresponding to a separation of $3.55 - 2.33 = 1.22$ Myr between the two extinction pulses. Similarly, the smallest value of $(t_{br} - t_{os})$ that is consistent with the observed data is found at the bottom right vertex of the joint confidence region and has coordinates $t_{os} = 2.45$ Myr and $t_{br} = 3.17$ Myr, corresponding to a separation of $3.17 - 2.45 = 0.72$ Myr. We therefore conclude that the ostracode and brachiopod extinction pulses were separated by at least 0.72 Myr and at most 1.22 Myr. In other words, the interval (0.72 Myr, 1.22 Myr) is a 95% confidence interval for the time separating the two extinction pulses. Note that this interval does not contain zero. Thus, we are confident that the extinction pulses were not simultaneous, which is necessarily consistent with the results of hypothesis test III.

In general, the 95% confidence interval for $(\theta_{br} - \theta_{os})$ has the form $(Lower, Upper)$, where the lower and upper endpoints are given by

$$Lower = \min(t_{br} - t_{os}) \text{ among points } (t_{os}, t_{br}) \\ \text{in the 95\% joint confidence region,}$$

and

$$Upper = \max(t_{br} - t_{os}) \text{ among points } (t_{os}, t_{br}) \\ \text{in the 95\% joint confidence region.} \quad (6)$$

This interval is conservative in that its probability of being correct is at least 95% for all values of θ_{os} and θ_{br} ; the actual coverage probability is usually slightly higher. This occurs because there may be data values that would lead us to reject the true values of θ_{os} and θ_{br} , but that would nonetheless yield an interval containing $(\theta_{br} - \theta_{os})$. Using the statistical software R, we ran simulations to determine the empirical coverage probability of our method. We found that intervals generated using equation (6) were correct (in that they contained the true difference $[\theta_{br} - \theta_{os}]$) in 96.7% of simulated random samples, so empirically the method does work (at least) 95% of the time. A formal proof of the correctness of the interval is given in Appendix 3.

Estimating the Time Separating Extinction Pulses: A Simple Formula

The endpoints of the confidence interval given in (6) can also be expressed as a simple algebraic formula. Let k and $(G - k)$ represent the number of ostracode and brachiopod genera, respectively, and let N_{os} and N_{br} represent the total number of fossil finds for these genera; that is, $N_{os} = \sum_{i=1}^k n_i$ and $N_{br} = \sum_{i=k+1}^G n_i$.

Note that the joint confidence region (Fig. 5) is bounded by three sides. Two of these are the straight lines $t_{os} = y_{os}$ and $t_{br} = y_{br}$, where y_{os} denotes the time of the latest fossil find among all ostracode genera and y_{br} is defined similarly for the brachiopod genera. For the Meishan data, $y_{os} = 2.33$ Myr and $y_{br} = 3.17$ Myr. Values to the left of $t_{os} = 2.33$ Myr or below $t_{br} = 3.17$ Myr are impossible, as they would result in genera going extinct before their latest fossil finds. The third side is given by the curve whose equation is

$$N_{os} \log(t_{os}) + N_{br} \log(t_{br}) = C, \quad (7)$$

where $C = \sum_{i=1}^G n_i \log(y_i) + \chi^2_{2G, .95}/2$ (see Appendix 4). This curve is shown as a dotted line in Figure 5. Points lying below this line (and within the region bounded by the lines $t_{os} = 2.33$ Myr and $t_{br} = 3.17$ Myr) are those for which the null hypothesis is not rejected, whereas points above this line are those for which the null hypothesis is rejected.

Using this information, we can express the confidence interval for $(\theta_{br} - \theta_{os})$ in terms of simple algebraic formulas. The lower endpoint of the confidence interval (corresponding to the bottom right corner of the joint confidence region) is the intersection of equation (7) and the line $t_{br} = y_{br}$. The upper endpoint (corresponding to the top left corner of the joint confidence region) is the intersection of equation (7) and the line $t_{os} = y_{os}$. Solving for these intersections (see Appendix 4 for details), we obtain the following formulas for the confidence interval for $(\theta_{br} - \theta_{os})$:

$$\begin{aligned} Lower &= y_{br} - e^{C/N_{os}} / y_{br}^{N_{br}/N_{os}} \quad \text{and} \\ Upper &= e^{C/N_{br}} / y_{os}^{N_{os}/N_{br}} - y_{os}. \end{aligned} \quad (8)$$

For the Meishan data, we have

$$\begin{aligned} C &= \sum_{i=1}^G n_i \log(y_i) + \chi^2_{2G, .95}/2 = 215.6 \\ N_{os} &= \sum_{i=1}^k n_i = 153 & N_{br} &= \sum_{i=k+1}^G n_i = 68 \\ y_{os} &= 2.33 & y_{br} &= 3.17. \end{aligned}$$

Substituting these values into (8), we find that $Lower = 0.72$ and $Upper = 1.22$. Thus the 95% confidence interval for $(\theta_{br} - \theta_{os})$ is (0.72 Myr, 1.22 Myr), which is the same as the interval we determined graphically in the previous section. Thus, based on their stratigraphic ranges, we are 95% confident that the ostracode extinction pulse and the brachiopod extinction pulse were separated by 720,000 to 1.22 million years.

Discussion

Most previous work on statistical methods for mass extinctions has focused on hypothesis tests and confidence intervals for simultaneous extinctions. For pulsed extinction events, however, no quantitative methods have been developed that account for the Signor-Lipps effect. We have described a method for estimating the time separating two extinction pulses in a pulsed extinction event that accounts for the incompleteness of the fossil record.

The Permian Meishan data set we used here was expressed in units of time. In the more common case of data expressed in units of stratigraphic distance, our method may be applied to compute a confidence interval for the distance separating the two extinction horizons. If the sedimentation rate is known and constant, it may be possible to convert this interval to one expressed in units of time.

Our data set dealt with the extinction of two taxonomic groups (ostracodes and brachiopods), but our methodology is applicable to other types of comparisons as well. For instance, if data are available, one could estimate the time separating the extinctions of groups of taxa having different ecological roles, life habits, larval strategies, or other characteristics. It is important to note that the two groups of taxa must be specified a priori according to

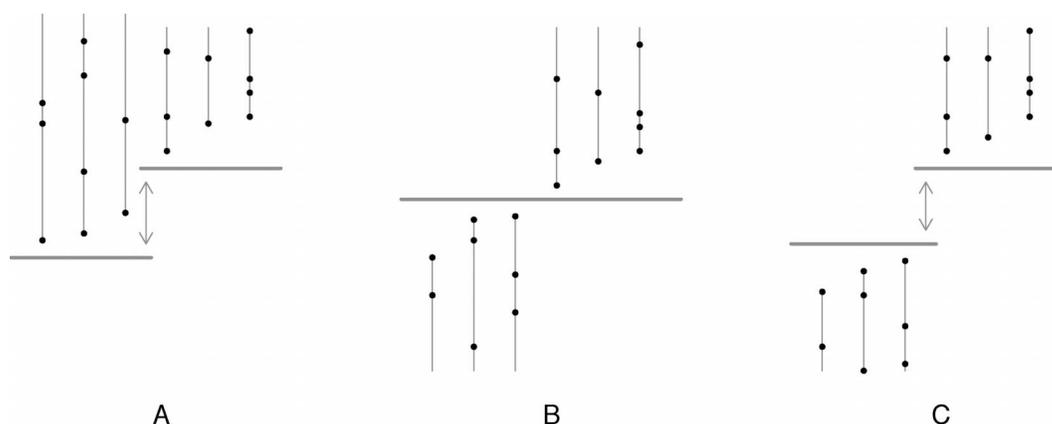


FIGURE 6. Examples of additional situations in which the methodology presented here can be applied. A, Estimating the time or stratigraphic distance separating two sets of simultaneous originations. Such a situation might arise in the study of recovery after a mass extinction. B, Testing whether a set of taxa go extinct simultaneously at the same time as a set of simultaneous originations. Such a situation might arise in the study of coordinated stasis. C, Estimating the time or stratigraphic distance separating a mass extinction from its recovery.

the traits of the fossil organisms, not according to their apparent extinction patterns.

An intriguing application of our method would be to estimate the time separating terrestrial and marine extinction pulses in a mass extinction. For example, Twitchett et al. (2001) present data from a Permo-Triassic boundary section in Greenland preserving both an abundant marine fauna and terrestrial palynomorphs. Such sections are rare, however, and typically multiple sections will need to be correlated and calibrated to place them on a single time or thickness scale. For such a resulting composite section, correlation methods exist that can estimate error bars on the relative position of fossil finds projected from different source sections (Sadler and Cooper 2003; Sadler 2004). Incorporating such error estimates into our methodology is a topic for further research.

Our method can also be “reversed” to compute confidence intervals for the time separating origination pulses, allowing tests of recovery patterns following mass extinctions (Payne et al. 2006). A simplified example is shown in Figure 6A. The method can also be applied to both extinction and origination in a single section, thus allowing tests for the clustering of first and last occurrences that characterizes coordinated stasis (Brett and Baird 1995) (Fig. 6B). The method can also be used to estimate the time separating an ex-

tinguishing event and its subsequent recovery (Wilf and Johnson 2004; Ward et al. 2005) (Fig. 6C).

Our model assumes continuous sampling with uniform preservation and recovery potential. In particular, the assumption of uniform preservation and recovery is a strong one. If recovery potential is in fact decreasing over time (as would be expected if populations were in decline), then range extensions will likely be underestimated, and the calculated confidence interval will be too narrow. On the other hand, if recovery potential is increasing over time (as would be expected if populations were growing), the calculated confidence interval will likely be overly conservative (i.e., wider than necessary). Future work will aim to relax the assumption of uniform recovery potential. To do so, it will be necessary to estimate recovery potential quantitatively, perhaps using additional information such as water depth (Holland 2003) or abundance. A fully analytical methodology for confidence intervals in such situations will likely be intractable; a practical solution will most likely require Monte Carlo simulation methods.

Acknowledgments

The authors thank S. Chang, K. Angielczyk, A. Bush, H. Chernoff, D. Erwin, L. Ivany, C. Marshall, J. Payne, P. Roopnarine, and M. Stein for their assistance; and G. Hunt, P. Sadler, and

P. Wagner for helpful reviews. Financial support from the Swarthmore College Research Fund, the Christian R. and Mary F. Lindback Foundation, and the Woodrow Wilson National Fellowship Foundation (S.C.W.) and the Eugene M. Lang Faculty Fellowship (P.J.E.) is gratefully acknowledged.

Literature Cited

- Alvarez, L. W., W. Alvarez, F. Asaro, and H. V. Michel. 1980. Extraterrestrial cause for the Cretaceous-Tertiary extinction. *Science* 208:1095–1108.
- Brett, C. E., and G. C. Baird. 1995. Coordinated stasis and evolutionary ecology of Silurian to Middle Devonian faunas in the Appalachian basin. Pp. 285–314 in D. H. Erwin and R. L. Anstey, eds. *New approaches to studying speciation in the fossil record*. Columbia University Press, New York.
- Casella, G., and R. L. Berger. 2002. *Statistical inference*, 2d ed. Duxbury, Pacific Grove, Calif.
- Foote, M. 1997. Estimating taxonomic durations and preservation probability. *Paleobiology* 23:278–300.
- Holland, S. M. 2003. Confidence limits on fossil ranges that account for facies changes. *Paleobiology* 29:468–479.
- Huelsenbeck, J. P., and B. Rannala. 1997. Maximum likelihood estimation of phylogeny using stratigraphic data. *Paleobiology* 23:174–180.
- Isozaki, Y. 2002. End-Permian double-phased mass extinction and killer volcanism. *Geological Society of America Abstracts with Programs* 34:360.
- Jin, Y. G., Y. Wang, W. Wang, Q. H. Shang, C. Q. Cao, and D. H. Erwin. 2000. Pattern of marine mass extinction near the Permian-Triassic boundary in South China. *Science* 289:432–436.
- Keller, G., W. Stinnesbeck, T. Adatte, and D. Stüben. 2003. Multiple impacts across the Cretaceous-Tertiary boundary. *Earth-Science Reviews* 62:327–363.
- Knoll, A. H., R. K. Bambach, D. E. Canfield, and J. P. Grotzinger. 1996. Comparative Earth history and late Permian mass extinction. *Science* 273:452–457.
- Macellari, C. E. 1986. Late Campanian-Maastrichtian ammonite fauna from Seymour Island (Antarctic Peninsula). *Journal of Paleontology* 60(Suppl.).
- Marshall, C. R. 1995. Distinguishing between sudden and gradual extinctions in the fossil record: predicting the position of the Cretaceous-Tertiary iridium anomaly using the ammonite fossil record on Seymour Island, Antarctica. *Geology* 23:731–734.
- Marshall, C. R., and P. D. Ward. 1996. Sudden and gradual molluscan extinctions in the latest Cretaceous in western European Tethys. *Science* 274:1360–1363.
- McConway, K. J., and H. J. Sims. 2004. A likelihood-based method for testing for nonstochastic variation of diversification rates in phylogenies. *Evolution* 58:12–13.
- McGhee, G. R., Jr. 2001. The 'multiple impacts hypothesis' for mass extinction: a comparison of the Late Devonian and the Eocene. *Palaeogeography, Palaeoclimatology, Palaeoecology* 176:47–58.
- Paul, C. R. C., M. A. Lamolda, S. F. Mitchell, M. R. Vaziri, A. Gorostidi, and J. D. Marshall. 1999. The Cenomanian-Turonian boundary at Eastbourne (Sussex, UK): a proposed European reference section. *Palaeogeography, Palaeoclimatology, Palaeoecology* 150:83–121.
- Payne, J. L., D. J. Lehrmann, J. Wei, and A. H. Knoll. 2006. The pattern and timing of biotic recovery from the end-Permian extinction on the Great Bank of Guizhou, Guizhou Province, China. *Palaios* 21:63–85.
- Sadler, P. M. 2004. Quantitative biostratigraphy—achieving finer resolution in global correlation. *Annual Review of Earth and Planetary Sciences* 32:187–213.
- Sadler, P. M., and R. A. Cooper. 2003. Best-fit intervals and consensus sequences. Pp. 49–94 in P. J. Harries, ed. *Approaches in high-resolution stratigraphic paleontology*. Kluwer Academic, Dordrecht.
- Signor, P. W., and J. H. Lipps. 1982. Sampling bias, gradual extinction patterns, and catastrophes in the fossil record. In L. T. Silver and P. H. Schultz, eds. *Geological implications of large asteroids and comets on the Earth*. Geological Society of America Special Paper 190:291–296.
- Solow, A. R. 1996. Tests and confidence intervals for a common upper endpoint in fossil taxa. *Paleobiology* 22:406–410.
- Solow, A. R., and W. K. Smith. 1997. On fossil preservation and the stratigraphic ranges of taxa. *Paleobiology* 23:271–277.
- . 2000. Testing for a mass extinction without selecting taxa. *Paleobiology* 26:647–650.
- Springer, M. S. 1990. The effect of random range truncations on patterns of evolution in the fossil record. *Paleobiology* 16:512–520.
- Stanley, S. M., and X. Yang. 1994. A double mass extinction at the end of the Paleozoic Era. *Science* 266:1340–1344.
- Strauss, D., and P. M. Sadler. 1989. Classical confidence intervals and Bayesian probability estimates for ends of local taxon ranges. *Mathematical Geology* 21:411–427.
- Twitchett, R. J., C. V. Looy, R. Morante, H. Visscher, and P. B. Wignall. 2001. Rapid and synchronous collapse of marine and terrestrial ecosystems during the end-Permian biotic crisis. *Geology* 29:351–354.
- Wagner, P. J. 2000. Likelihood tests of hypothesized durations: determining and accommodating biasing factors. *Paleobiology* 26:431–449.
- Wang, S. C., and C. R. Marshall. 2004. Improved confidence intervals for estimating the position of a mass extinction boundary. *Paleobiology* 30:5–18.
- Ward, P. D., J. Botha, R. Buick, M. O. de Kock, D. H. Erwin, G. H. Garrison, J. L. Kirschvink, and R. Smith. 2005. Abrupt and gradual extinction among Late Permian land vertebrates in the Karoo Basin, South Africa. *Science* 307:709–713.
- Wilf, P., and K. R. Johnson. 2004. Land plant extinction at the end of the Cretaceous: a quantitative analysis of the North Dakota megafossil record. *Paleobiology* 30:347–368.
- Xie, S., R. D. Pancost, H. Yin, H. Wang, and R. P. Evershed. 2005. Two episodes of microbial change coupled with Permo/Triassic faunal mass extinction. *Nature* 434:494–497.

Appendix 1

Calculating the Likelihood Ratio Statistic

Here we provide a detailed derivation of the likelihood ratio test statistic. Recall that for genus i , θ_i denotes the true extinction time and t_i denotes the hypothesized extinction time. The number of fossil finds is denoted by n_i , the time of the j th fossil find is denoted by x_{ij} , and the vector of all fossil finds for genus i is denoted by $x_i = \{x_{i1}, x_{i2}, \dots, x_{in_i}\}$.

For an individual genus i , conditional on n_i , it follows from the assumption of uniform preservation and recovery that the random variable X_{ij} is uniformly distributed over the interval $[0, \theta_i]$. The probability density function (pdf) for a single fossil find X_{ij} is thus $f(x) = 1/\theta_i$ if $x \leq \theta_i$ (otherwise $f(x) = 0$ if $x > \theta_i$). Then the joint pdf for a random sample of n_i finds is $f(x_{i1}, \dots, x_{in_i}) = 1/\theta_i^{n_i}$ if $x_{ij} \leq \theta_i$ for all j , and the joint pdf for all G taxa (assuming their mutual independence) is $f(x_1, \dots, x_G) = \prod_{i=1}^G 1/\theta_i^{n_i}$ if $x_{ij} \leq \theta_i$ for all i and j .

The joint pdf is a function of the observed data, conditional on the hypothesized times. The likelihood function is numeri-

cally identical but viewed as a function of the hypothesized times, conditional on the observed data. We denote the likelihood function by $L(t_1, \dots, t_G | x_1, \dots, x_G)$, where t_1, \dots, t_G denote any set of hypothesized extinction times. Given a data set x_1, \dots, x_G ,

the likelihood function tells us how plausible it is that the hypothesized extinction times t_1, \dots, t_G were the ones that generated the observed data x_1, \dots, x_G .

The likelihood ratio statistic, denoted by λ , is equal to

$$\lambda = \frac{\text{Value of } L(t_1 \dots t_G | x_1 \dots x_G) \text{ using the hypothesized times } t_1 \dots t_G \text{ specified by } H_0}{\text{Maximum value } L(t_1 \dots t_G | x_1 \dots x_G) \text{ can attain using any possible values of } t_1 \dots t_G}.$$

The numerator of this ratio is the value of the likelihood function evaluated at the times t_1, \dots, t_G hypothesized by the null hypothesis. The denominator is the maximum possible value of the likelihood function. This maximum value is attained, by definition, when the likelihood function is evaluated at the maximum likelihood estimates (MLEs) of $\theta_1, \dots, \theta_G$ —the values of t_1, \dots, t_G that make the observed data most probable. Using standard statistical theory, it is straightforward to show that the MLE for each θ_i is equal to the time of the latest fossil find for that genus, denoted by $y_i = \max(x_{ij})$. This gives us the following expression for the likelihood ratio statistic:

$$\lambda = \frac{\prod_{i=1}^G 1/t_i^{n_i}}{\prod_{i=1}^G 1/y_i^{n_i}} \quad \text{if } t_i \geq y_i \text{ for all taxa } i.$$

Simplifying, we arrive at $\lambda = \prod_{i=1}^G (y_i/t_i)^{n_i}$ if $t_i \geq y_i$ for all taxa i ($\lambda = 0$ otherwise), which is identical to equation (1) above.

Appendix 2

The Sampling Distribution of $-2 \log \lambda$

Here we show that the sampling distribution of $-2 \log \lambda$, twice the negative logarithm of the likelihood ratio statistic, has a chi-square distribution with $2G$ degrees of freedom, where G denotes the number of genera. As given earlier, $-2 \log \lambda$ is defined as follows:

$$-2 \log \lambda = -2 \sum_{i=1}^G n_i \log \left(\frac{Y_i}{t_i} \right) \quad \text{if } t_i \geq Y_i \text{ for all taxa } i.$$

Under the null hypothesis we have $\theta_i = t_i$, so the X_{ij} are uniformly distributed over the interval $[0, t_i]$. It follows that the ratio X_{ij}/t_i is uniformly distributed over the interval $[0, 1]$. As a result, Y_i/t_i has the same distribution as the largest order statistic of a sample of size n_i from a uniform(0, 1) distribution, which can be shown by standard theory to be a Beta($n_i, 1$) random variable. Let $B_i = Y_i/t_i$, and let $W_i = -\log(B_i)$. We use the cumulative distribution function (cdf) method to derive the sampling distribution of W_i . The cdf of W_i is defined as $F(w) = P(W_i \leq w)$. This can be rewritten as follows:

$$\begin{aligned} F(w) &= P(W_i \leq w) \\ &= P[-\log(B_i) \leq w] \quad (\text{sub in definition of } W_i) \\ &= P(B_i \geq e^{-w}) \quad (\text{take negatives and exponentiate}) \\ &= \int_{e^{-w}}^1 n_i b^{n_i-1} db \quad (\text{use the Beta pdf}) \\ &= b^{n_i} \Big|_{e^{-w}}^1 \quad (\text{evaluate the integral}) \\ &= 1 - e^{-n_i w} \quad (\text{substitute}) \end{aligned}$$

To find the probability density function (pdf) of W_i , $f(w)$, we take the derivative of the cdf found above:

$$\begin{aligned} f(w) &= \frac{d}{dw} F(w) = \frac{d}{dw} (1 - e^{-n_i w}) \quad (\text{substitute } F(w) \text{ from above}) \\ &= n_i e^{-n_i w} \quad \text{for } w > 0 \end{aligned}$$

This last expression is the pdf of an exponential distribution with rate parameter n_i , so W_i has an exponential(n_i) distribution. It then follows from standard statistical theory that the product $2n_i W_i$ has an exponential(1/2) distribution, which is equivalent to a chi-square(2) distribution. Because $-2 \log \lambda = \sum_{i=1}^G 2n_i W_i$, we

have that $-2 \log \lambda$ is a sum of G independent chi-square(2) random variables, and therefore it has a chi-square($2G$) distribution.

Appendix 3

Derivation of the Confidence Interval for Estimating the Time Separating Extinction Pulses

Here we show that our method for estimating the time between extinction pulses yields a confidence interval with the desired coverage probability (e.g., 95%). Let X denote the observed data, consisting of the times of the fossil finds x_1, \dots, x_G . Consider testing the hypothesis

$$\begin{aligned} H_0: \quad \theta_1 = \theta_2 = \dots = \theta_{21} = t_{\text{os}}; \\ \theta_{22} = \theta_{23} = \dots = \theta_{34} = t_{\text{br}} \end{aligned} \quad (5)$$

for a set of hypothesized extinction times $(t_{\text{os}}, t_{\text{br}})$. Given the hypothesized times $(t_{\text{os}}, t_{\text{br}})$, let $R(t_{\text{os}}, t_{\text{br}}) = \{X : \text{Reject } H_0(t_{\text{os}}, t_{\text{br}}) \text{ at } X\}$ be the rejection region, the set of all observed data sets X for which the null hypothesis is rejected. Let $R^c(t_{\text{os}}, t_{\text{br}})$ be the complement of the rejection region, the set of all X for which the null hypothesis is not rejected (i.e., the set of all X that are consistent with the null hypothesis).

For a given data set X , let $A(X) = \{(t_{\text{os}}, t_{\text{br}}) : X \in R^c(t_{\text{os}}, t_{\text{br}})\}$, the set of all points $(t_{\text{os}}, t_{\text{br}})$ for which X is consistent with extinctions at times t_{os} and t_{br} . Define $Lower(X) = \min_{(t_{\text{os}}, t_{\text{br}}) \in A(X)} (t_{\text{br}} - t_{\text{os}})$, the smallest value of $t_{\text{br}} - t_{\text{os}}$ separating extinction times that are consistent with the observed data X . Similarly, define $Upper(X) = \max_{(t_{\text{os}}, t_{\text{br}}) \in A(X)} (t_{\text{br}} - t_{\text{os}})$, the largest value of $t_{\text{br}} - t_{\text{os}}$ separating extinction times that are consistent with the observed data X .

Let $(t_{\text{os}}^*, t_{\text{br}}^*)$ be the actual extinction times. Consider $R^c(t_{\text{os}}^*, t_{\text{br}}^*)$, the set of data sets X that are consistent with the actual extinction times. For an arbitrary data set X , if X is in $R^c(t_{\text{os}}^*, t_{\text{br}}^*)$, then the null hypothesis

$$\begin{aligned} H_0: \quad \theta_1 = \theta_2 = \dots = \theta_{21} = t_{\text{os}}^*; \\ \theta_{22} = \theta_{23} = \dots = \theta_{34} = t_{\text{br}}^* \end{aligned}$$

would not be rejected (at level α) for X . This is equivalent to saying that the point $(t_{\text{os}}^*, t_{\text{br}}^*)$ is in $A(X)$. As a result, $Lower(X)$ must be at least as small as $(t_{\text{br}}^* - t_{\text{os}}^*)$; similarly, $Upper(X)$ must be at least as large as $(t_{\text{br}}^* - t_{\text{os}}^*)$. This then implies that $Lower(X) \leq (t_{\text{br}}^* - t_{\text{os}}^*) \leq Upper(X)$. In other words, the interval $(Lower(X), Upper(X))$ contains the true difference $(t_{\text{br}}^* - t_{\text{os}}^*)$.

Thus, if a data set X is in $R^c(t_{\text{os}}^*, t_{\text{br}}^*)$, then the interval $(Lower(X), Upper(X))$ contains the true difference $(t_{\text{br}}^* - t_{\text{os}}^*)$. But the probability that an arbitrary data set X is in $R^c(t_{\text{os}}^*, t_{\text{br}}^*)$ is .95, which is a consequence of the alpha level being .05. Therefore, $(Lower(X), Upper(X))$ has probability .95 of covering $(t_{\text{br}}^* - t_{\text{os}}^*)$, so it is a 95% confidence interval.

Note that intervals constructed in this way are often conser-

vative—that is, their true coverage probability may be higher than 95%. Why does this occur? For the actual extinction times (t_{os}^* , t_{br}^*), there is a 95% chance of getting a sample X such that (t_{os}^* , t_{br}^*) is not rejected. This means that the difference $t_{br}^* - t_{os}^*$ is in the confidence interval ($Lower(X)$, $Upper(X)$) for these 95% of possible samples. However, there may be other samples for which (t_{os}^* , t_{br}^*) is rejected, but for which the difference $t_{br}^* - t_{os}^*$ nonetheless falls in the confidence interval generated by that sample. This increases the coverage probability, and thus the interval ($Lower(X)$, $Upper(X)$) has a coverage probability of at least 95%.

Appendix 4

Algebraic Formulas for the Confidence Interval

Here we derive the simple algebraic formulas for the confidence interval for ($\theta_{br} - \theta_{os}$) given in equation (8).

Recall from equation (2) that the test statistic is

$$-2 \log \lambda = -2 \sum_{i=1}^G n_i \log \left(\frac{y_i}{t_i} \right) \text{ if } t_i \geq y_i \text{ for all taxa } i. \quad (2)$$

Using properties of logarithms, this can be rewritten as

$$-2 \log \lambda = -2 \sum_{i=1}^G n_i \log(y_i) + 2 \sum_{i=1}^G n_i \log(t_i).$$

We reject the null hypothesis $H_0: \theta_1 = t_1, \theta_2 = t_2, \dots, \theta_G = t_G$ when $-2 \log \lambda > \chi_{2G, 95}^2$, the 95th percentile of a chi-square distribution with $2G$ degrees of freedom. After rearranging terms and dividing by 2, this is equivalent to rejecting H_0 when

$$\sum_{i=1}^G n_i \log(t_i) > \sum_{i=1}^G n_i \log(y_i) + \chi_{2G, 95}^2 / 2. \quad (9)$$

Note that for a two-pulse extinction, $t_1 = t_2 = \dots = t_k$ and $t_{k+1} = t_{k+2} = \dots = t_G$, with $t_k \neq t_{k+1}$. It then follows that (9) is equivalent to rejecting H_0 when

$$N_{os} \log(t_{os}) + N_{br} \log(t_{br}) > \sum_{i=1}^G n_i \log(y_i) + \chi_{2G, 95}^2 / 2.$$

For convenience, let C denote the right-hand side of this inequality. This implies that the region in Figure 5 for which we reject H_0 is bounded below by the equation

$$N_{os} \log(t_{os}) + N_{br} \log(t_{br}) = C \quad (7)$$

as well as by the lines $t_{os} = y_{os}$ and $t_{br} = y_{br}$, where y_{os} is the time of the latest fossil find among all ostracode genera, and y_{br} is the time of the latest fossil find among all brachiopod genera.

The lower endpoint of the confidence interval for ($\theta_{br} - \theta_{os}$) is determined by the bottom right corner of the joint confidence region, which gives the smallest difference $t_{br} - t_{os}$ among extinction pulses that are consistent with the observed data. This point is the intersection of (7) with the line $t_{br} = y_{br}$, where $y_{br} = 3.17$ Myr in our example. The vertical coordinate of this point is $t_{br} = y_{br}$; its horizontal coordinate is found by substituting in $t_{br} = y_{br}$ and solving for t_{os} in (7) as follows:

$$\begin{aligned} N_{os} \log(t_{os}) + N_{br} \log(y_{br}) &= C \\ \log(t_{os}) &= [C - N_{br} \log(y_{br})] / N_{os} \\ t_{os} &= \exp\{[C - N_{br} \log(y_{br})] / N_{os}\} \\ t_{os} &= \exp(C / N_{os}) / \exp\{[N_{br} / N_{os}] \log(y_{br})\} \\ t_{os} &= e^{C / N_{os}} / y_{br}^{N_{br} / N_{os}}. \end{aligned}$$

We therefore find that the coordinates for the bottom right corner of the joint confidence region are ($e^{C / N_{os}} / y_{br}^{N_{br} / N_{os}}$, y_{br}), and it is at this point that the difference $t_{br} - t_{os}$ is minimized among extinction pulses that are consistent with the data. This difference is $t_{br} - t_{os} = y_{br} - e^{C / N_{os}} / y_{br}^{N_{br} / N_{os}}$, which is the lower endpoint for the confidence interval given in (8).

Similarly, the upper endpoint is determined by the top left corner of the joint confidence region, which gives the largest difference $t_{br} - t_{os}$ among extinction pulses that are consistent with the observed data. This is the intersection of (7) with the line $t_{os} = y_{os}$, where $y_{os} = 2.33$ Myr in our example. This point has horizontal coordinate $t_{os} = y_{os}$; its vertical coordinate is found by substituting in $t_{os} = y_{os}$ and solving for t_{br} in (7) in a manner similar to that above, which gives

$$t_{br} = e^{C / N_{br}} / y_{os}^{N_{os} / N_{br}}.$$

We therefore find that the coordinates for the top left corner of the joint confidence region are (y_{os} , $e^{C / N_{br}} / y_{os}^{N_{os} / N_{br}}$), and it is at this point that the difference $t_{br} - t_{os}$ is maximized among extinction pulses that are consistent with the data. This difference is $t_{br} - t_{os} = e^{C / N_{br}} / y_{os}^{N_{os} / N_{br}} - y_{os}$, which is the upper endpoint for the confidence interval given in (8).