

## Optimal estimators of the position of a mass extinction when recovery potential is uniform

Steve C. Wang, David J. Chudzicki, and Philip J. Everson

*Abstract.*—Numerous methods have been developed to estimate the position of a mass extinction boundary while accounting for the incompleteness of the fossil record. Here we describe the point estimator and confidence interval for the extinction that are optimal under the assumption of uniform preservation and recovery potential, and independence among taxa. First, one should pool the data from all taxa into one combined “supersample.” Next, one can then apply methods proposed by Strauss and Sadler (1989) for a single taxon. This gives the optimal point estimator in the sense that it has the smallest variance among all possible unbiased estimators. The corresponding confidence interval is optimal in the sense that it has the shortest average width among all possible intervals that are invariant to measurement scale. These optimality properties hold even among methods that have not yet been discovered. Using simulations, we show that the optimal estimators substantially improve upon the performance of other existing methods. Because the assumptions of uniform recovery and independence among taxa are strong ones, it is important to assess to what extent they are satisfied by the data. We demonstrate the use of probability plots for this purpose. Finally, we use simulations to explore the sensitivity of the optimal point estimator and confidence interval to nonuniformity and lack of independence, and we compare their performance under these conditions with existing methods. We find that nonuniformity strongly biases the point estimators for all methods studied, inflates their standard errors, and degrades the coverage probabilities of confidence intervals. Lack of independence has less effect on the accuracy of point estimates as long as recovery potential is uniform, but it, too, inflates the standard errors and degrades confidence interval coverage probabilities.

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

\*Present address: Department of Mathematics, University of Chicago, 5734 University Avenue, Chicago, Illinois 60637. E-mail: djc@math.uchicago.edu

Accepted: 2 February 2009

### Introduction

Given a collection of taxa going extinct in a single or composite stratigraphic section, we often want to know whether they went extinct simultaneously, and if so, at what time or stratigraphic position (Jin et al. 2000; Pearson et al. 2001; Barry et al. 2002; Wilf and Johnson 2004; Ward et al. 2005; Wilson 2005; Groves et al. 2007). These goals, however, are complicated by the incompleteness of the fossil record (Signor and Lipps 1982). Many authors have addressed the question of how to distinguish simultaneous versus gradual extinction in the fossil record (Strauss and Sadler 1989; Springer 1990; Meldahl 1990; Marshall 1995; Solow 1996; Marshall and Ward 1996; Solow and Smith 2000; Wagner 2000; Payne 2003; Wang and Marshall 2004; Solow et al. 2006; Wang and Everson 2007). If the locations of fossil finds are consistent with a

simultaneous extinction, we can then use existing methods for estimating the time or position of the extinction boundary (Strauss and Sadler 1989; Springer 1990; Marshall 1995; Solow 1996; Marshall and Ward 1996; Solow and Smith 2000; Wang and Marshall 2004). Because several such methods exist, it is of interest to determine which method works best in a given situation.

In this paper, we derive the optimal point estimator and optimal confidence interval under the assumptions of (1) uniform preservation and recovery potential, and (2) independence among taxa. By optimal, we mean that these methods are the most precise estimators possible under reasonable conditions. Specifically, we consider a point estimator to be optimal if it has the smallest possible variance or standard error among all possible unbiased estimators. Similarly, we consider a method for finding confidence

intervals to be optimal if it yields the shortest intervals among all possible scale-invariant methods. Using statistical theory, we show that these optimal methods involve combining all taxa in the section into one “super-sample” and then applying the methods of Strauss and Sadler (1989) for finding point estimates and confidence intervals. We then discuss how to check the assumptions of uniformity and independence, and we explore the sensitivity of our methods and competing methods to violations of these assumptions.

### Properties and Assumptions

Suppose we have range charts giving locations of fossil finds for a collection of taxa in a single or composite stratigraphic section. We assume that a hypothesis test has already shown that the taxa could have gone extinct simultaneously (Strauss and Sadler 1989; Springer 1990; Meldahl 1990; Marshall 1995; Solow 1996; Marshall and Ward 1996; Huelssenbeck and Rannala 1997; Solow and Smith 2000; Wagner 2000; Wang and Marshall 2004; Solow et al. 2006; Wang and Everson 2007); such tests are not discussed further here. Several methods have been developed for estimating the time or stratigraphic position of the simultaneous extinction boundary (Strauss and Sadler 1989; Springer 1990; Marshall 1995; Solow 1996; Marshall and Ward 1996; Solow and Smith 2000; Wang and Marshall 2004); henceforth we will refer only to position for convenience. A natural question is which of these methods is preferable or optimal in a given situation.

Previous work (Wang and Marshall 2004) used simulations to compare six methods for calculating confidence intervals on the position of an extinction boundary. However, to show that a method is optimal in a more general sense—even among methods that have not yet been invented—it is necessary to construct a mathematical proof that demonstrates its optimality under some specified set of conditions. We find it useful to divide these conditions into two types: *properties* that we would like the method to possess, and *assumptions* that we make about the structure of the data. Here we specify two desired

properties of the method. First, we want our point estimator to be *unbiased*—that is, it should on average give the correct answer, rather than producing estimates that are systematically too high or too low. Second, we want our confidence interval to be *scale invariant*—that is, it should give consistent results regardless of the choice of measurement scale. For instance, scale invariance guarantees that our results do not differ depending on whether we record fossil locations in feet or meters, which would be nonsensical. (Although we do not explicitly require this property of our point estimator, it turns out that the optimal point estimator will be scale invariant as well.) We also specify two assumptions about the nature of the data: (1) *uniformity*, that the probability of a fossil being preserved and recovered throughout the section is constant, and (2) *independence*, that the positions of any taxon’s fossil finds are independent of those of any other taxon’s. Under these conditions, we will derive the optimal point estimator and confidence interval for the position of the extinction boundary. In other words, we will show that our method is optimal among all possible methods that possess these specified properties, when applied to data following these assumptions.

### Assumptions: Uniform Recovery and Independence

Let  $k$  denote the number of taxa in the section. We assume that each taxon is extant at the base of the section, which we assign position 0, and that all taxa go extinct simultaneously at an unknown position up-section denoted by  $\theta$ . Our goal is to derive optimal estimators of  $\theta$ . Let  $n_i$  denote the number of fossil finds or horizons for taxon  $i$ .

We assume that preservation and recovery potential is uniform over the true range of each taxon. That is, for each taxon  $i$ , the positions of its fossil finds or horizons are uniformly distributed over its true range  $[0, \theta]$ . We further assume that the positions of fossil finds for any taxon are independent of those for any other taxon. That is, knowledge of the positions of fossil finds of one taxon tells us nothing about the fossil finds of any

other taxon. An additional implicit assumption is that the section is sampled continuously. If this is not the case, positions of fossil finds will be correlated for different taxa, because they will tend to cluster at the discrete sampling points. For our purposes, we will subsume this assumption into that of independence because their effects are similar. Although these assumptions are strong ones, they are commonly made—either explicitly or implicitly, and not always with full justification—in existing literature (Paul 1982; Macellari 1986; Springer and Lilje 1988; Strauss and Sadler 1989; Springer 1990; Marshall 1990, 1995; Marshall and Ward 1996; Solow 1996; Jin et al. 2000; Solow and Smith 2000; Solow et al. 2006; Wang and Everson 2007).

Assuming these assumptions are satisfied, a key insight is that the positions of the fossil finds for all taxa should be pooled into one “supersample” and treated as if they came from a single taxon. That is, we may treat the  $k$  samples from taxa 1, 2, ...,  $k$ , having sizes  $n_1, n_2, \dots, n_k$ , as one supersample of size  $N$ , where  $N = n_1 + n_2 + \dots + n_k$ . The positions of all  $N$  fossil finds will then be uniformly distributed over the range  $[0, \theta]$ . This follows from the assumption that the fossil finds are independently and uniformly distributed over  $[0, \theta]$  for each taxon, because a combination of independent uniformly distributed samples is itself uniformly distributed. To see why this is true, consider an analogy. Suppose one person picks three random integers between 0 and 100—say, 11, 34, and 66—and another person independently picks two random integers between 0 and 100—say, 53 and 87. If we combine the random integers chosen by both people—11, 34, 53, 66, and 87—the result is equivalent to one sample of five randomly chosen integers between 0 and 100.

Thus, combining independent uniformly distributed samples, even those having different sample sizes, yields a uniformly distributed supersample, and the two situations are equivalent. We have therefore reduced our multiple-taxon scenario to a pooled single-taxon scenario, and we now need only to find the optimal estimators for this latter scenario. The equivalence of these two sce-

narios is critical to our argument, because any method that is optimal for the pooled single-taxon scenario must also be optimal for the multiple-taxon scenario.

It may seem counterintuitive from a paleontological viewpoint to combine data from different taxa—perhaps representing organisms as different as corals, ammonites, and mosasaurs—but doing so is a necessary consequence of the assumptions of uniformity and independence. This approach differs from that of previous works (Springer 1990; Marshall 1995; Marshall and Ward 1996; Solow 1996; Wang and Marshall 2004), all of which treat each taxon as a distinct entity.

### Deriving the Optimal Estimators

Henceforth we will consider our data to be a single supersample consisting of the pooled data for all taxa. We now show that the well-known methods given by Strauss and Sadler (1989) for a single taxon are the optimal point estimator and confidence interval for such a pooled data set.

Let  $X$  denote the highest fossil find among all taxa. Given our assumptions of uniformity and independence, it is straightforward to show that  $X$  is a sufficient statistic—that is,  $X$  contains all the information that exists in the data about  $\theta$  (Casella and Berger 2002: p. 277). Following standard statistical practice, we therefore consider only estimators that are functions of  $X$ .

First, we describe a derivation of the optimal point estimator for  $\theta$ . Because this is a well-known proof that can be found in many mathematical statistics textbooks, we give only a brief outline and refer the reader to the references for details. We will restrict our attention to unbiased estimators of  $\theta$ —that is, estimators that are, on average, “on target.” Clearly  $X$  itself is not unbiased, as  $X$  almost always underestimates  $\theta$ . However, the estimator  $(\frac{N+1}{N})X$  is an unbiased estimator of  $\theta$  (Casella and Berger 2002: p. 230), and because it is based on a complete sufficient statistic, it follows that  $(\frac{N+1}{N})X$  is the minimum variance unbiased estimator of  $\theta$  (Casella and Berger 2002: p. 346). Therefore,  $(\frac{N+1}{N})X$  is the optimal estimator of  $\theta$ , in the sense that it has the smallest possible variance

among all unbiased estimators of  $\theta$ . This fact is noted by Strauss and Sadler (1989) and Solow (2005), but in the context of a single taxon. Note that even though we did not impose scale invariance in this proof, the optimal estimator nonetheless possesses this desirable property.

We now turn our attention to the optimal confidence interval for  $\theta$  with confidence level  $(1 - \alpha)100\%$ . (Here  $\alpha$  denotes the alpha level of a corresponding hypothesis test. For instance, to achieve 95% confidence we take  $\alpha = 0.05$  so that  $(1 - \alpha)100\% = (1 - 0.05)100\% = 95\%$ ). To our knowledge, this derivation is new in the literature. We present the derivation in three steps. As before, we consider only functions of the sufficient statistic  $X$ , the highest fossil find among all taxa. We further restrict consideration to methods that are scale invariant. Intuitively, we should expect the same results regardless of the units in which our data are recorded, whether in feet, meters, or any other units; the position of the estimated extinction should refer to the same physical location in the rock record. This has the consequence of restricting the class of possible  $(1 - \alpha)100\%$  confidence intervals to those having the form  $(aX, bX)$ , where  $a$  and  $b$  are constants:

*Step 1:* A scale-invariant confidence interval must have the form  $(aX, bX)$ .

*Proof:* By definition, a function  $f(x)$  is scale invariant if  $f(kx) = k f(x)$  for all  $k$  and all  $x$ . Let  $k = 1$ ; then  $f(x) = f(1x) = f(x1) = x f(1)$ . This shows that a scale-invariant function  $f(x)$  has the form  $cx$  for some constant  $c$ , where  $c = f(1)$ . Because the endpoints of the confidence interval must be functions of the sufficient statistic  $X$ , they must have the form  $(aX, bX)$  for some constants  $a$  and  $b$ .

Now we determine the values of  $a$  and  $b$ . We begin by deriving the relationship between  $a$  and  $b$ :

*Step 2:* The constant  $b$  can be written in terms of the constant  $a$  as follows:  $b = a [1 - a^N(1 - \alpha)]^{-1/N}$ .

*Proof:* Let  $Y = X/\theta$ . Standard statistical theory shows that  $Y$  has a Beta( $N, 1$ )

distribution with cumulative distribution function  $F(y) = P(Y \leq y) = y^N$  for  $0 \leq y \leq 1$  (Casella and Berger 2002: p. 230). To achieve a confidence level of  $(1 - \alpha)100\%$ , we require that  $P(aX \leq \theta \leq bX) = 1 - \alpha$ . The left-hand side of this equality can be simplified as follows:

$$\begin{aligned} P(aX \leq \theta \leq bX) &= P(a \leq \theta/X \leq b) \\ &= P(1/b \leq X/\theta \leq 1/a) \\ &= P(1/b \leq Y \leq 1/a) \\ &= P(Y \leq 1/a) \\ &\quad - P(Y \leq 1/b) \\ &= F(1/a) - F(1/b) \\ &= a^{-N} - b^{-N}. \end{aligned}$$

Thus we have  $a^{-N} - b^{-N} = 1 - \alpha$ . We now solve for  $b$ :

$$\begin{aligned} b^{-N} &= a^{-N} - (1 - \alpha) \\ b &= [a^{-N} - (1 - \alpha)]^{-1/N} \quad (1) \\ &= a [1 - a^N(1 - \alpha)]^{-1/N}. \end{aligned}$$

This proves Step 2.

Note that in equation (1),  $a^N(1 - \alpha) \leq 1$  must hold in order to avoid taking an even root of a negative quantity, which constrains  $a$  such that  $a \leq (1 - \alpha)^{-1/N}$ . Also note that  $a \geq 1$  must hold, which is a consequence of  $X \leq \theta$ .

Finally, we solve for  $a$  and  $b$ :

*Step 3:* The shortest  $(1 - \alpha)100\%$  confidence interval of the form  $(aX, bX)$  has  $a = 1$  and  $b = \alpha^{-1/N}$ .

*Proof:* For any  $X$ , the width of the confidence interval is  $(bX - aX) = (b - a)X$ . This quantity is minimized when  $(b - a)$  is minimized. Using equation (1), we have

$$\begin{aligned} (b - a) &= a [1 - a^N(1 - \alpha)]^{-1/N} - a \\ &= a \left( [1 - a^N(1 - \alpha)]^{-1/N} - 1 \right). \end{aligned}$$

The expression  $[1 - a^N(1 - \alpha)]^{-1/N}$  increases as  $a$  increases within the range from  $a = 1$  to  $a = (1 - \alpha)^{-1/N}$ . Thus the width of the interval is minimized by choosing  $a$  as small as possible; hence  $a = 1$ .

To solve for  $b$ , we set  $a = 1$  in equation (1), which gives us  $b = 1[1 - 1^N(1 - \alpha)]^{-1/N} = [1 - (1 - \alpha)]^{-1/N} = \alpha^{-1/N}$ . This proves Step 3.

We have thus shown that the optimal  $(1 - \alpha)100\%$  confidence interval, having the shortest width for any  $X$ , is  $(X, X\alpha^{-1/N})$ , where  $X$  is the highest fossil find in the section. This is the same interval given by Strauss and Sadler (1989), although they give a different derivation and estimate both the lower and upper endpoints of the stratigraphic range instead of assuming a fixed base. A third derivation is given by Gingerich and Uhen (1998). However, neither of these references proves or claims optimality of the interval.

### Performance of the Optimal Estimators: Simulation Results

The preceding proof shows that the methods of Strauss and Sadler (1989), applied to a pooled supersample of taxa, give the optimal point estimator and confidence interval. However, it does not demonstrate how much better these methods perform compared to competing methods. To investigate their performance, we ran simulations with 10,000 randomly generated data sets having ten taxa each. For each taxon, the number of fossil horizons was randomly generated from a Poisson distribution having mean = 6, constrained to have a minimum of two horizons and a maximum of 20 horizons. A simultaneous extinction was simulated to lie at  $\theta = 100$  m above the base of the section. For each data set, we estimated the position of the simultaneous extinction boundary using the optimal point estimator and confidence interval described above and four competing methods: those of Springer (1990), Marshall (1995), Solow (1996), and Wang and Marshall (2004). For the methods of Marshall (1995) and Wang and Marshall (2004), we used the midpoint of the confidence interval as the point estimate. For the method of Springer (1990), we used the position at which the  $p$ -value of the Kolmogorov-Smirnov test was maximized (i.e., the position at which the data were most consistent with a simultaneous extinction). Solow (1996) did not describe a method for finding a point estimate; we used a 50% confidence interval constructed using

his method as the point estimate. For confidence intervals using each of the five methods, we used a confidence level of 89%. (Because of discreteness in the number of taxa, the methods of Marshall [1995] and Wang and Marshall [2004] are able to determine confidence intervals only for certain confidence levels; thus we used 89% rather than 90%, which would not have been possible with our data sets.)

The left column of Figure 1A displays the sampling distributions of the 10,000 point estimates for each method and gives their mean and standard error. The middle column displays 100 randomly selected confidence intervals for each method and gives the empirical coverage level for all 10,000 simulated data sets (i.e., the percent of confidence intervals out of the 10,000 that correctly contained the true value of  $\theta = 100$  m). The right column displays the widths of the 10,000 intervals and gives the mean width for each method. As can be seen from the left column, all five methods gave point estimates that were approximately unbiased (within  $\pm 3\%$  of the correct answer), but the optimal method was the most accurate, with a mean estimate of 100.001 m. It was also the most precise, as expected, with a standard error (1.67 m) slightly smaller than that of Wang and Marshall (2004) and substantially smaller than those of the other three methods. In addition, the optimal point estimate was correct within  $\pm 5\%$  (i.e., between 95 m and 105 m) in 98.2% of simulated samples, compared to 63.6% for Marshall (1995), 77.3% for Solow (1996), 88.7% for Springer (1990), and 97.2% for Wang and Marshall (2004) (not shown in figure).

As can be seen from the middle and right columns, all five methods gave confidence intervals with empirical coverage probabilities close to the nominal value of 89%, but they varied substantially in their mean width (= upper endpoint - lower endpoint). The optimal confidence interval is typically somewhat shorter than intervals given by Wang and Marshall (2004) and two to five times shorter than those given by the other three methods, allowing us to pinpoint the position of an extinction boundary with the greatest possible precision.

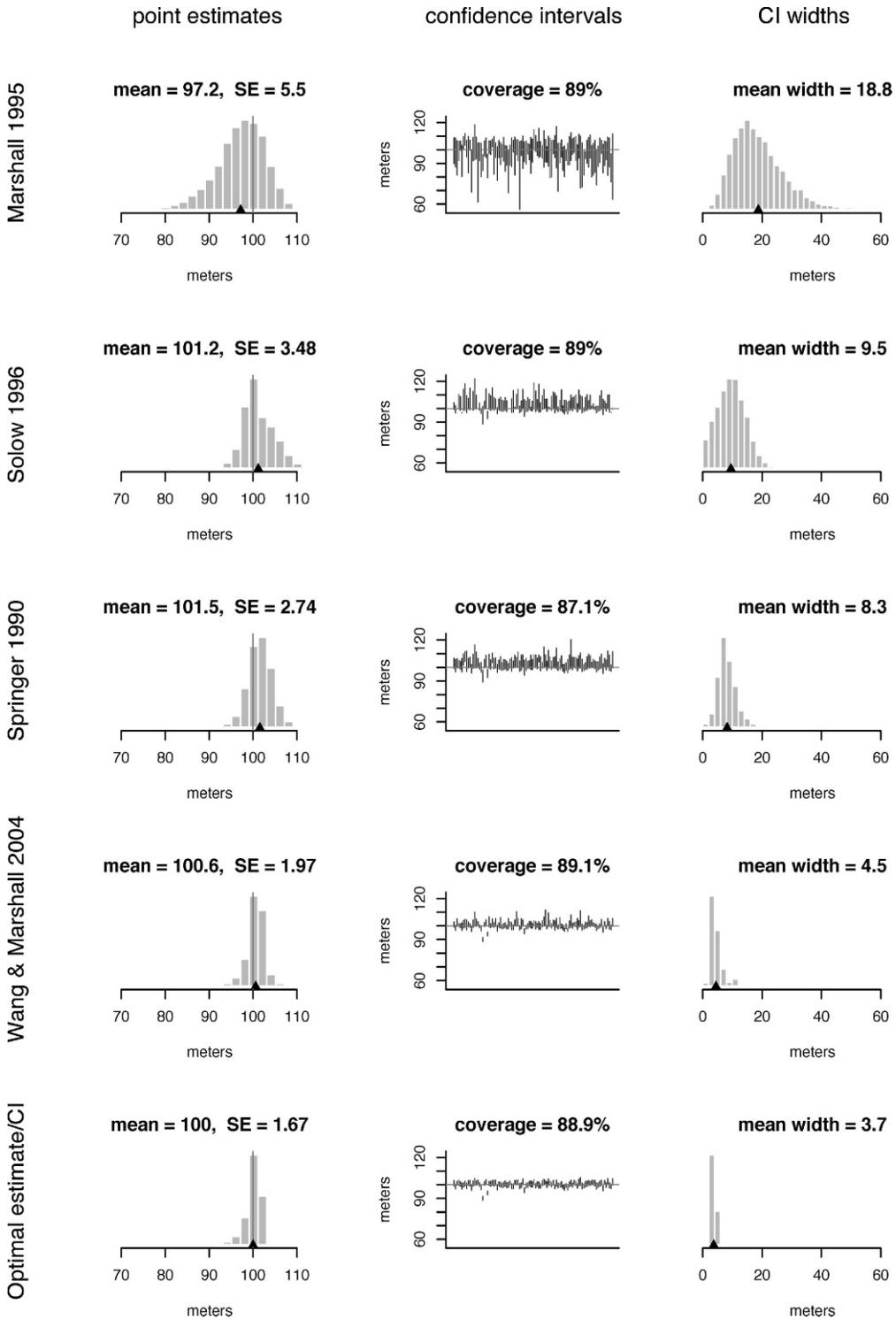


FIGURE 1. Results from 10,000 simulated data sets. The true extinction was set at  $\theta = 100$ . For each data set, we applied the methods proposed by Marshall (1995), Solow (1996), Springer (1990), and Wang and Marshall (2004), and the optimal method described here, which uses the estimator of Strauss and Sadler (1989) applied to a pooled "supersample." **Left column:** Histograms of point estimates, with means and standard errors of the estimates. Vertical line denotes the true position of the extinction; triangle indicates the mean of the 10,000 estimates for each method. The

Note that the methods of both Springer (1990) and Solow (1996) are based on inverting a hypothesis test—that is, their  $(1 - \alpha)100\%$  confidence intervals consist of the set of positions not rejected by a corresponding hypothesis test at level  $\alpha$ . Some simulated data sets were not consistent with simultaneous extinction at *any* position (using  $\alpha = 1 - 0.89 = 0.11$ ), and thus the Springer (1990) and Solow (1996) methods were unable to calculate a confidence interval for these data sets. This occurred in 0.5% of data sets for the former method and 6.25% of data sets for the latter. For these data sets, we considered the Springer (1990) and Solow (1996) methods to have given incorrect confidence intervals. In addition, we excluded these data sets when calculating the mean point estimate, standard error, and mean confidence interval width for these two methods. However, these data sets are included in the results for the other three methods, which do not rely on inverting hypothesis tests and are thus always able to calculate a point estimate and confidence interval. (Results were essentially identical regardless of whether these data sets were excluded or included for the other three methods.)

### Checking Assumptions

The assumptions of uniform preservation and recovery and independence among taxa are strong ones and should not be accepted blindly. In fact, like an ideal gas in chemistry or a frictionless plane in physics, uniformity and independence are idealized scenarios; in

reality, they can never be fully satisfied. Therefore, what is important is not deciding *whether* or not they are violated—they are always false—but rather exploring *how* they are violated, and the effects of such violations. It may be the case that certain types of violations are more severe, while others may be more tolerable.

The issue of scale is relevant here. Because of facies control on fossil occurrences and sequence stratigraphic controls on facies changes in outcrops, it is inevitable that taxa in a stratigraphic section will have some degree of nonuniformity in recovery potential when analyzed at a sufficiently high level of resolution (Holland 2000, 2003; Holland and Patzkowsky 2002). For example, Holland (2003: Fig. 4) calculated recovery potential for the trilobite *Cryptolithus* and the brachiopod *Sowerbyella* of the Upper Ordovician Kope Formation near Cincinnati. In the recovery functions of both taxa, there is little secular trend in recovery potential at the level of some of the 20-meter sedimentary cycles (e.g., the cycle labeled C1-2 in the figure). At this gross scale of resolution, it may be reasonable to consider a uniform approximation to the recovery function. However, at a higher level of resolution, strong high-frequency variation corresponding to meter-scale sedimentary cycles is clearly visible in his figure.

There are also reasons to believe that the assumption of independence may often not be met. Clusters of fossil occurrences are expected at flooding surfaces and at sequence

---

←

optimal method has the smallest standard error of any possible unbiased estimator. It was also the most accurate of the five methods (mean = 100.001 m, no significant difference from 100 m). The means of the other four methods were significantly different from 100 m, although they were close, averaging within  $\pm 3\%$  of the correct answer. **Middle column:** 100 randomly selected confidence intervals calculated using each method; empirical coverage probabilities are also given (i.e., the percentage of confidence intervals that included the true position of the extinction [horizontal line]). Coverage probabilities for all methods were not significantly different from the nominal 89% confidence level except for that of Springer (1990), which was still close to the nominal level (off by 1.9 percentage points). **Right column:** Histograms of widths of 89% confidence intervals (upper endpoint – lower endpoint). Triangle indicates the mean of the 10,000 widths for each method. The optimal method has the shortest average width of any possible method that is invariant to measurement scale. Note that the Springer (1990) and Solow (1996) methods are not able to compute a valid confidence interval for some data sets (see text for details). This happened in 0.5% of data sets for the former method and 6.25% of data sets for the latter. We counted these methods as having given incorrect confidence intervals for such data sets. These data sets were excluded when calculating the mean point estimate, standard error, and mean confidence interval width for these two methods, but not for the other three methods.

boundaries, especially when a large proportion of taxa are stenotopic (Holland 1995; Holland and Patzkowsky 2002). Furthermore, we have assumed that sampling is continuous; under discrete sampling, fossil occurrences will be confined to individual beds or bedding planes, also causing clustering among taxa (Marshall 1994; Payne 2003). It is crucial, then, to assess the extent to which the assumptions of uniformity and independence are violated, and in what manner.

To check the assumption of uniformity, Marshall (1990) used Kolmogorov-Smirnov tests, and Solow et al. (2006) and Vogel et al. (2009) used uniform probability plots. Here we expand upon the approach of the last two references in checking the assumptions of both uniformity and independence. Probability plots (sometimes referred to as theoretical quantile-quantile plots [Chambers et al. 1983; Solow et al. 2006; Vogel et al. 2009]) are a powerful graphical tool for assessing whether a data set follows a specified distribution. These plots display the ordered observed values on the y-axis, plotted against the ordered expected values (under the assumed distribution) on the x-axis. For example, in a uniform probability plot, we would plot the following  $n$  points:

- (1) the minimum observed value versus the expected minimum in a sample of size  $n$  from a uniform distribution,
- (2) the second-smallest observed value versus the expected second-smallest value in a sample of size  $n$  from a uniform distribution,
- ...
- ( $n$ ) the maximum observed value versus the expected maximum in a sample of size  $n$  from a uniform distribution.

If it is plausible that the data set is sampled from a uniform distribution, then the  $n$  plotted points should lie close to the line  $y = x$ . Confidence bands can be added to the plot to indicate the expected range of random variation of points around the line. For example, Figure 2A shows a uniform probability plot for a data set on radiocarbon ages of Alaskan mammoth finds, redrawn from Solow et al. (2006). All the points fall close to the line  $y = x$  and within the confidence

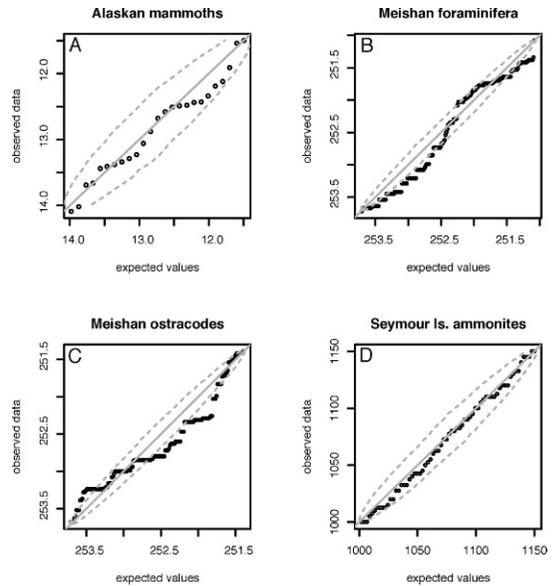


FIGURE 2. Examples of probability plots for assessing uniformity and independence. The y-axis represents the actual positions of fossil finds in the data set; the x-axis represents the expected positions if recovery potential were uniform. If recovery potential is in fact uniform, then the data points should lie close to the  $y = x$  line (solid gray line). Dashed gray lines represent 95% confidence limits. A, Alaskan mammoth data, redrawn from Figure 2A of Solow et al. (2006). No apparent deviations from uniformity are evident. Units are radiocarbon ages (Ka). Note that we have plotted points instead of line segments as in the original figure, and we have ignored dating error for simplicity. B, Meishan foraminifera data from Jin et al. (2000). Deviations from uniformity are apparent: the points do not all fall within the confidence limits, especially in the top right of the plot, and evidence for lack of independence can be seen (horizontal “stripes” in lower half of plot). Units are inferred ages (Ma). C, Meishan ostracode data from Jin et al. (2000), also analyzed in Wang and Everson (2007). Although deviations from uniformity and lack of independence are apparent, points in the top right of the plot are approximately consistent with a uniform expectation. Thus, although uniformity and independence are not satisfied, this data set may still provide useful information for some analyses, such as inferring the common extinction time of these taxa. Units are inferred ages (Ma). D, Upper 150 m of Seymour Island (Antarctica) ammonite data from Macellari (1995) and analyzed in Strauss and Sadler (1989), Springer (1990), Marshall (1995), Solow (1996), and Solow and Smith (2000). Only fossil finds higher than 1000 m above the base of the section are shown, as this is the part of the section when all ten taxa are likely to be extant. There is slight evidence for lack of independence (horizontal stripes), but overall the positions of fossil finds are consistent with uniform recovery potential. Units are meters above the base of the section.

bands, so the assumptions of uniform recovery and independence appear reasonable.

Figure 2B shows a uniform probability plot of the Meishan foraminifera data from Jin

et al. (2000). In contrast to Figure 2A, the points do not all lie within the confidence bounds (which are very narrow because of the large sample size), showing that the uniform model may not be appropriate. This is particularly true of points in the top right of the plot, which is especially problematic because such points carry the most information about the time of extinction. Furthermore, evidence of non-independence can be seen in the bottom left corner of the plot. The points tend to lie along horizontal "stripes," because in this dataset the recorded locations of fossil finds were keyed to individual beds (D. H. Erwin personal communication 2008). (This can also be seen directly in Figure 2B of Jin et al. 2000.) Furthermore, note that most of the highest finds (above  $y = 251.7$  Ma) tend to be lower (older) than would be expected under a uniform model. If we were to apply methods assuming uniformity and independence to this data set, then, we would likely overestimate the age of the true extinction time.

Figure 2C shows a uniform probability plot of the Meishan ostracode data from Jin et al. (2000), which were also analyzed by Wang and Everson (2007). As in Figure 2B, the points do not all lie within the confidence bounds, and there is evidence for non-independence (horizontal stripes near  $y = 253.24, 253.0, 252.85,$  and  $252.34$  Ma). However, if we are interested primarily in the taxon's extinction time, then these data may provide a reasonable estimate despite their lack of uniformity and independence. The points in the top right of the graph (above  $y = 251.8$  Ma) lie within the confidence bands and show little evidence of non-independence, and it is these points that are the most important in estimating the time of extinction. Thus, although this data set as a whole violates the assumptions of uniformity and independence, it may nonetheless provide useful information for testing some kinds of hypotheses.

As a final example, we consider the Seymour Island, Antarctica, ammonite data from Macellari (1986), commonly used as an example of data having uniform recovery (Strauss and Sadler 1989; Springer 1990;

Marshall 1995; Solow 1996; Solow and Smith 2000). A uniform probability plot of the entire data set (not shown) reveals that the uniform assumption is grossly violated, primarily because two taxa are found several hundred meters lower than most of the other taxa. However, if we limit the data set to only the fossil finds occurring above 1000 m above the base of the section (the range over which all taxa are likely extant), then the data do in fact closely follow a uniform recovery function (Fig. 2D). There may be some slight correlation among taxa (e.g., at  $y = 1012.5$  and  $1110$  m), but overall the data appear to be "close enough" to uniformity and independence to be treated as such. (A likelihood-based method for identifying regions of constant recovery within a taxon's stratigraphic range is given by Wagner [2000]).

Note that it is possible to use a probability plot to carry out a formal hypothesis test (for example, by using the correlation coefficient of the points as a test statistic [Filliben 1975; Vogel et al. 2009]). However, we have chosen not to take this approach, because we believe the primary benefit of a probability plot is that it gives detailed information on *how* the data deviate from the assumed distribution and to what extent, rather than just a binary "reject"/"do not reject" decision. As discussed above, even data sets that deviate significantly from uniformity may still hold valuable information for some analyses.

### Sensitivity to Assumptions

In this section we explore the sensitivity of our estimators to common violations of the assumptions of uniformity and independence. Previous work in the context of a single taxon has shown that the point estimate and confidence interval of Strauss and Sadler (1989) are adversely affected by a nonuniform recovery function (Solow 2003), but to our knowledge the effect of nonindependence among taxa on point estimates and confidence intervals has not been studied quantitatively.

We simulated continuously sampled data sets from a variety of scenarios, including recovery functions that were (1) uniform, (2) linearly decreasing up-section, or (3) nonli-

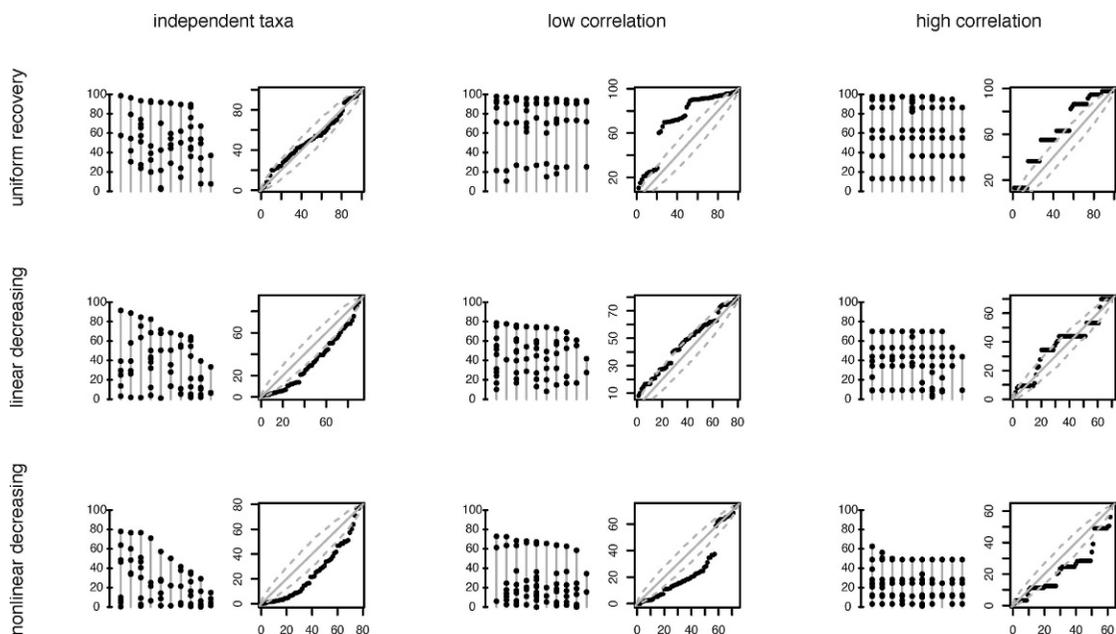


FIGURE 3. Examples of simulated data sets. Each data set has ten taxa; the number of fossil finds for each taxon was randomly generated from a Poisson distribution having mean = 6, constrained to have a minimum of two finds and a maximum of 20 finds. A simultaneous extinction was simulated to lie at  $\theta = 100$  m above the base of the section. For each data set, a range chart and a uniform probability plot are shown. Data were simulated from nine different combinations of recovery potential and independence: uniform (top row), linearly decreasing (middle), and nonlinearly decreasing (bottom) recovery potential, and independent (left column), weakly correlated (middle), and strongly correlated (right) finds among taxa. Nonlinear recovery potential followed a Beta(1,3) probability density, a decreasing function in the shape of a hollow decay curve.

nearly decreasing up-section. That is, as we approach the true time of extinction of the taxon, its fossils become harder to recover, perhaps because of sea level variation, declining abundance, decreased sampling effort, taphonomic changes, or other reasons. (Of course, it is also possible that recovery potential is increasing up-section. However, we chose to simulate decreasing recovery potential in our data sets because that scenario is the “hard” case compared to a uniform recovery function, whereas increasing recovery potential (e.g., due to increased sampling up-section approaching a mass extinction event) is the “easier” case because the position of the extinction is better constrained.) We also simulated data sets with taxa whose fossil finds were (1) independent, (2) weakly correlated, or (3) strongly correlated with those of other taxa, to simulate the effect of fossils being confined to discrete beds or bedding planes. This gave us a total of nine different combinations of recovery functions and correlations. Figure 3 shows exam-

ples of data sets randomly generated under each of these nine scenarios.

For each scenario, we simulated 10,000 data sets, each having ten taxa. The number of fossil horizons for each taxon was randomly generated from a Poisson distribution with mean = 6, constrained to have a minimum of two horizons and a maximum of 20 horizons. The locations of the horizons were then placed randomly according to the specified recovery function and correlation structure. A simultaneous extinction was simulated to lie at  $\theta = 100$  m above the base of the section. For each simulated data set, we calculated the optimal point estimate and 89% confidence interval, as well as estimates using the four methods described earlier. The results are summarized in Figure 4.

*Violations of Uniformity.*—As was seen in Figure 1, Figure 4 shows that when the recovery function was uniform and taxa were independent, the point estimator was approximately unbiased for all five methods, and confidence intervals achieved their nominal

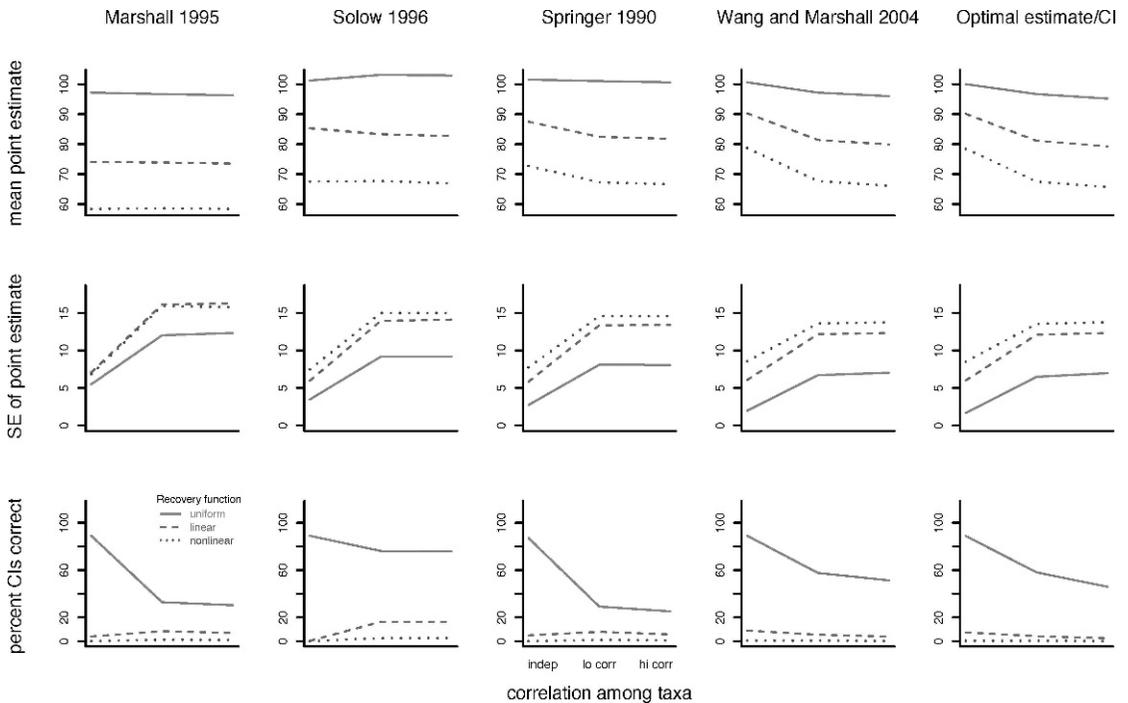


FIGURE 4. Results from simulations assessing sensitivity to nonuniform recovery potential and lack of independence among taxa. Solid lines represent uniform recovery potential; dashed lines represent linearly declining recovery; dotted lines represent nonlinearly declining recovery. Within each panel, taxa are simulated independently (left), with weak correlation (center), and with strong correlation (right). See Figure 3 caption for details on the simulations. **Top row:** Means of point estimates using the five methods tested; the true position of the simultaneous extinction was set at  $\theta = 100$  m above the base of the section. **Middle row:** Standard errors of point estimates. **Bottom row:** Coverage probabilities for 89% confidence intervals. Nonuniformity in the recovery function severely degrades results for all five methods: point estimates are 10–40% too low, standard errors roughly double, and 89% confidence intervals achieve only 0–20% coverage. Lack of independence has relatively little effect on point estimates when recovery is uniform, but standard errors are roughly doubled and confidence interval coverage probabilities are decreased.

coverage probabilities (solid lines). Under nonuniform recovery functions, however, the results were dramatically worse. When the recovery function was linearly declining up-section (dashed lines), point estimates for all methods were 10–25% too low, standard errors were roughly doubled, and coverage probabilities for 89% confidence intervals were under 20% for all methods. That is, instead of being correct for 89% of data sets, confidence intervals were in fact incorrect 80% to nearly 100% of the time. Performance was even worse when the recovery function was nonlinearly declining (dotted lines): point estimates for all methods were 20–40% too low, standard errors were roughly doubled, and coverage probabilities for 89% confidence intervals ranged from 0% to 3% for all methods. In summary, all of these methods are very sensitive to the assumption

of uniform recovery, and it is therefore crucial to check this assumption.

*Violations of Independence.*—When recovery potential was uniform, lack of independence in the positions of fossil finds among taxa had relatively little effect on the point estimates for all five methods (Fig. 4, solid lines in top row of plots). However, lack of independence caused the point estimates to be biased downward for three of the methods when recovery potential was not uniform (Fig. 4, dashed and dotted lines in top row). Lack of independence also caused the standard errors to roughly double regardless of the recovery function (Fig. 4, middle row). When the recovery function was uniform, lack of independence caused the coverage probabilities of the confidence intervals to drop substantially for all methods except that of Solow (1996) (Fig. 4, solid lines in bottom

row). For nonuniform recovery functions, however, coverage probabilities were already so low that lack of independence did little to lower them further (Fig. 4, dashed and dotted lines in bottom row). In summary, for all of these methods point estimates are approximately unbiased in the presence of correlations among taxa as long as recovery potential is uniform, but standard errors increase substantially, and coverage probabilities are degraded as well. It is therefore crucial to check the assumption of independence.

### Summary and Discussion

Given a collection of taxa in a single or composite stratigraphic section thought to have gone extinct simultaneously, and assuming uniform recovery and independence among taxa, one should first pool the positions of their finds together into a single combined "supersample." One can then apply the methods proposed by Strauss and Sadler (1989) for estimating the true extinction of a single taxon. We have proven that these methods are optimal in the sense that they are the most precise estimators satisfying the properties of unbiasedness and scale invariance, under the assumption that the data follow uniform recovery potential and independence among taxa. This optimality applies not only among the methods tested in this paper, but among all possible methods satisfying the stated conditions—even ones that have not yet been invented.

Our proof depends on the assumptions of uniform recovery and independence among taxa, and it is important to assess the reasonableness of these assumptions. We advocate the use of probability plots to explore the extent to which these assumptions are met or violated, rather than a hypothesis-testing approach. Using simulations, we assessed the sensitivity of the optimal estimators and competing methods to violations of uniformity and independence. Nonuniformity had particularly adverse effects, strongly biasing point estimates and inflating their standard errors, and degrading coverage probabilities of confidence intervals. Lack of independence had less effect on the accuracy of point estimates, but it also caused inflated

standard errors and degraded coverage probabilities. Clearly, checking these assumptions is crucial in any analysis of this kind.

When recovery potential is not uniform, none of the existing methods are appropriate, and modifying the optimal methods is not straightforward. In such cases, it may be possible to adapt the methods of Springer (1990) and Wang and Marshall (2004). These methods require estimating range extensions for individual taxa, for which one can apply generalized estimates that account for changing recovery potential (Marshall 1997; Solow 2003). Judging from Figure 1, we expect that in such situations point estimates and confidence intervals based on Wang and Marshall (2004) will be substantially more precise than those based on Springer (1990).

We have assumed that all taxa are initially extant at some position 0, corresponding to the base of the section or some point higher in the section, rather than originating at different positions within the section. This appears to be a reasonable assumption for at least some data sets (e.g., Jin et al. 2000: Fig. 2A). This assumption is necessary to justify pooling data from different taxa. It should be possible to derive optimal methods under a more general model with different origination times; this is an area of ongoing research.

We derived the optimal point estimator while requiring the property of unbiasedness—that the estimator is neither systematically high nor systematically low. It may be possible to derive an estimator with smaller variance if we do not insist on unbiasedness. However, such an approach is likely to be analytically intractable. Furthermore, as the optimal method already has a relative error of less than 2%, it is unlikely that any other method, unbiased or not, would improve on it substantially.

### Acknowledgments

We thank S. Chang, D. Erwin, C. Groot, C. Marshall, H. Shapiro, and P. Wagner for their assistance, and J. Payne and S. Holland for their thoughtful reviews. Acknowledgment is made to the Donors of the American Chemical Society Petroleum Research Fund for partial support of this research (PRF#

44152-GB8 to PI Wang). Funding from the Swarthmore College Research Fund is also gratefully acknowledged.

### Literature Cited

- Barry, J. C., M. E. Morgan, L. J. Flynn, D. Pilbeam, A. K. Behrensmeyer, S. M. Raza, I. A. Khan, C. Badgley, J. Hicks, and J. Kelley. 2002. Faunal and environmental change in the Late Miocene Siwaliks of northern Pakistan. *Paleobiology Memoir* 4. *Paleobiology* 28(Suppl. to No. 2).
- Casella, G., and R. L. Berger. 2002. *Statistical inference*, 2d ed. Duxbury, Pacific Grove, Calif.
- Chambers, J. M., W. S. Cleveland, B. Kleiner, and P. A. Tukey. 1983. *Graphical methods for data analysis*. Wadsworth, Pacific Grove, Calif.
- Filliben, J. J. 1975. The probability plot correlation coefficient test for normality. *Technometrics* 17:111–117.
- Gingerich, P. D., and M. D. Uhen. 1998. Likelihood estimation of the time of origin of Cetacea and the time of divergence of Cetacea and Artiodactyla. *Palaentologia Electronica* 1:45.
- Groves, J. R., R. Rettori, J. L. Payne, M. D. Boyce, D. Altiner. 2007. End-Permian mass extinction of Lagenide foraminifers in the southern Alps (northern Italy). *Journal of Paleontology* 81:415–434.
- Holland, S. M. 1995. The stratigraphic distribution of fossils. *Paleobiology* 21:92–109.
- . 2000. The quality of the fossil record: a sequence stratigraphic perspective. In D. H. Erwin and S. L. Wing, eds. *Deep time: Paleobiology's perspective*. *Paleobiology* 26(Suppl. to No. 4):148–168.
- . 2003. Confidence limits on fossil ranges that account for facies changes. *Paleobiology* 29:468–479.
- Holland, S. M., and M. E. Patzkowsky. 2002. Stratigraphic variation in the timing of first and last occurrences. *Palaio* 17:134–146.
- Huelsenbeck, J. P., and B. Rannala. 1997. Maximum likelihood estimation of phylogeny using stratigraphic data. *Paleobiology* 23:174–180.
- 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.
- Macellari, C. E. 1986. Late Campanian-Maastrichtian ammonite fauna from Seymour Island (Antarctic Peninsula). *Paleontological Society Memoir* 18. *Journal of Paleontology* 60(Suppl. to No. 2).
- Marshall, C. R. 1990. Confidence intervals on stratigraphic ranges. *Paleobiology* 16:1–10.
- . 1994. Confidence intervals on stratigraphic ranges: partial relaxation of the assumption of a random distribution of fossil horizons. *Paleobiology* 20:459–469.
- . 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.
- . 1997. Confidence intervals on stratigraphic ranges with nonrandom distributions of fossil horizons. *Paleobiology* 23:165–173.
- 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.
- Meldahl, K. H. 1990. Sampling, species abundance, and the stratigraphic signature of mass extinction: a test using Holocene tidal flat molluscs. *Geology* 18:890–893.
- Paul, C. R. C. 1982. The adequacy of the fossil record. In K. A. Joysey and A. E. Friday, eds. *Problems of phylogenetic reconstruction*. Systematics Association Special Volume 21:75–117. Academic Press, London.
- Payne, J. L. 2003. Applicability and resolving power of statistical tests for instantaneous extinction events in the fossil record. *Paleobiology* 29:37–51.
- Pearson, D. A., T. Schaefer, K. R. Johnson, and D. J. Nichols. 2001. Palynologically calibrated vertebrate record from North Dakota consistent with abrupt dinosaur extinction at the Cretaceous-Tertiary boundary. *Geology* 29:36–39.
- 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.
- . 2003. Estimation of stratigraphic ranges when fossil finds are not randomly distributed. *Paleobiology* 29:181–185.
- . 2005. Inferring extinction from a sighting record. *Mathematical Biosciences* 195:47–55.
- Solow, A. R., and W. K. Smith. 2000. Testing for a mass extinction without selecting taxa. *Paleobiology* 26:647–650.
- Solow, A. R., D. L. Roberts, and K. M. Robbirt. 2006. On the Pleistocene extinctions of Alaskan mammoths and horses. *Proceedings of the National Academy of Sciences USA* 103:7351–7353.
- Springer, M. S. 1990. The effect of random range truncations on patterns of evolution in the fossil record. *Paleobiology* 16:512–520.
- Springer, M. S., and A. Lilje. 1988. Biostratigraphy and gap analysis: the expected sequence of biostratigraphic events. *Journal of Geology* 96:228–236.
- 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.
- Vogel, R. M., J. R. M. Hosking, C. S. Elphick, D. L. Roberts, and J. M. Reed. 2009. Goodness of fit of probability distributions for sightings as species approach extinction. *Bulletin of Mathematical Biology* 71:701–719.
- Wagner, P. J. 2000. Likelihood tests of hypothesized durations: determining and accommodating biasing factors. *Paleobiology* 26:431–449.
- Wang, S. C., and P. J. Everson. 2007. Confidence intervals for pulsed mass extinction events. *Paleobiology* 33:324–336.
- 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.
- Wilson, G. P. 2005. Mammalian faunal dynamics during the last 1.8 million years of the Cretaceous in Garfield County, Montana. *Journal of Mammalian Evolution* 12:53–76.