Ecology, 90(5), 2009, pp. 1291–1300

The use of sighting records to infer species extinctions: an evaluation of different methods

Marcelo M. Rivadeneira,[1,3] Gene Hunt,[2] and Kaustuv Roy[1]

[1]*Section of Ecology, Behavior and Evolution, Division of Biological Sciences, University of California–San Diego, La Jolla, California 92093-0116 USA*

[2]*Department of Paleobiology, National Museum of Natural History, Smithsonian Institution, P.O. Box 37012 NHB MRC 121, Washington, D.C. 20013-7012 USA*

Abstract. In the absence of long-term monitoring data, inferences about extinctions of species and populations are generally based on past observations about the presence of a particular species at specified places and times (sightings). Several methods have been developed to estimate the probability and timing of extinctions from records of such sightings, but they differ in their computational complexity and assumptions about the nature of the sighting record. Here we use simulations to evaluate the performance of seven methods proposed to estimate the upper confidence limit on extinction times under different extinction and sampling scenarios. Our results show that the ability of existing methods to correctly estimate the timing of extinctions varies with the type of extinction (sudden vs. gradual) and the nature of sampling effort over time. When the probability of sampling a species declines over time, many of the methods perform poorly. On the other hand, the simulation results also suggest that as long as the choice of the method is determined by the nature of the underlying sighting data, existing methods should provide reliable inferences about the timing of past extinctions.

Key words: extinction probability; extinction time; false extinctions; local extinctions; sighting record; sudden vs. gradual extinction.

Introduction

With a few notable exceptions, global extinctions of species or local extinctions of populations are not recorded directly but have to be inferred from records of past occurrences or sighting records (Solow and Roberts 2003). For some species, information about past occurrences can be compiled from museum collections (Shaffer et al. 1998), whereas for others such as the dodo they may simply be reports of sightings (Roberts and Solow 2003). Reliably estimating the timing of extinctions of species (defined here as the definitive disappearance of a species either locally [local extinction] or globally) remains a major challenge. Because sighting records and museum collections are inherently incomplete, the absence of a species in such records does not necessarily mean that it is extinct. Instead it could reflect reduced sampling effort or short-term variation in the abundance of the species. As a result, the last time that a species is observed or sampled will generally occur sometime before its true extinction.

Manuscript received 14 February 2008; revised 2 July 2008; accepted 25 August 2008. Corresponding Editor (ad hoc): A. R. Solow.

[3] Present address: Centro de Estudios Avanzados en Zonas Aridas (CEAZA) and Departamento de Biología Marina, Facultad de Ciencias del Mar, Universidad Católica del Norte, Larrondo 1281, Coquimbo 1781421 Chile. E-mail: mrivadeneira@ucn.cl

Probabilistic methods can generate confidence intervals on the true extinction timing, given the sighting record, which can then be used to evaluate whether a species that has not been sighted for some time is likely to be truly extinct (Solow 1993*a*, 2005, Burgman et al. 1995, Roberts 2006).

Different methods have been proposed in the last two decades to generate confidence intervals on the timing of extinctions. These methods differ in their computational complexity and in the stringency of their assumptions about the nature of sampling. The simplest methods developed in the ecological literature (Solow 1993*a*, McInerny et al. 2006) are straightforward in their calculation, but require the restrictive assumption that sampling effort is constant over time. Other methods make no assumption about sampling intensities (Roberts and Solow 2003, Solow and Roberts 2003), but they tend to produce large confidence intervals (Solow 2005). A different set of methods explicitly model sampling effort in calculating confidence intervals and can weight absences depending on the intensity of sampling during the intervals in which the taxon was not observed (Burgman et al. 1995, McCarthy 1998, McInerny et al. 2006). Similar approaches also have been developed in paleontology (Strauss and Sadler 1989, Marshall 1997, Hayek and Bura 2001) as a way to estimate the real stratigraphic ranges of species and higher taxa from the inherently incomplete sampling of taxon occurrences in the fossil record (Marshall and Ward 1996, Labandeira et al. 2002, Wang and Marshall 2004, Solow et al. 2006*b*). The basic statistical properties of some of these methods have been explored in detail elsewhere (Strauss and Sadler 1989, Solow 1993*b*, Burgman et al. 1995, Grogan and Boreman 1998, McCarthy 1998, McInerny et al. 2006). In addition, Burgman et al. (2000) analyzed the performance of two of these methods under different scenarios of population decline and sampling intensities. Despite this work, it is not yet clear how the performance of each of the seven methods varies under different scenarios of extinction (e.g., sudden vs. gradual) and sampling. This is particularly relevant given the growing interest in reliably detecting local extinctions (Hedenas et al. 2002, Ferraz et al. 2003, Solow and Roberts 2003, Dulvy et al. 2004, Farnsworth and Ogurcak 2006, McInerny et al. 2006, Robbirt et al. 2006), and given the need to distinguish true extinctions from the failure to sample a species even when it is present (i.e., "Lazarus species," sensu Jablonski 1986, Roberts 2006).

In this study we focus on the performance of seven methods proposed in the ecological and paleontological literature that use sighting records or dated museum collections to estimate analytically the upper bound of the confidence interval (CI) of the timing of extinction of a species. The concept of estimating the confidence interval for extinction timing is rooted in the paleontological literature (Strauss and Sadler 1989, Marshall 1990, 1994, 1997, Marshall and Ward 1996, Wang and Marshall 2004), whereas many of the methods developed in the ecological literature focus on estimating extinction probabilities (Solow 1993*a*, *b*, McCarthy 1998, McInerny et al. 2006). The two approaches are very closely related (the upper confidence limit is the time at which the extinction probability reaches the nominal 0.05 value) and the usefulness of the confidence interval methods for ecological data has recently been highlighted (Roberts and Solow 2003, Solow and Roberts 2003). These approaches do, however, offer slightly different views of the same issue; extinction probabilities denote the plausibility of a taxon being extinct at a particular point in time, whereas CIs delimit the plausible temporal window for an extinction event.

A different class of models for estimating site occupancy and extinctions using survey data also exists in the literature and has been used to explore temporal changes in the compositions of ecological communities (Nichols et al. 1998, MacKenzie et al. 2003, 2006) as well as fossil assemblages (Nichols and Pollock 1983, Connolly and Miller 2001). The performance of some of these models under different sampling protocols has been evaluated recently (Dorazio 2007) and our aim here is to provide similar evaluations for models that are useful for the types of records available from museum collections and nonquantitative surveys.

Using simulations, we assessed each method under a set of different but realistic extinction and sampling scenarios. Confidence intervals produced by each method were evaluated based on their statistical coverage, which measures the degree to which intervals are appropriate in size. For an accurate and precise method, 95% of simulated extinctions should fall within the 95% confidence intervals. Our results show that under some conditions there can be pronounced differences in the performance of these methods, and we use these findings to provide some basic guidelines for their applications to real ecological and paleontological data sets.

Methods

Simulated data set

Artificial data sets of sightings were generated using different combinations of population dynamics and sampling effort. The length of the time series of sightings ranged uniformly from 20 to 120 time units (years), in order to mimic the length of time series typical of museum and herbarium collections. Then different sighting series were generated according to a variety of trends in the probability of occurrence or occupancy (P_o) and the probability of sampling (P_s) over the length of the series. We use the latter parameter as a measure of sampling effort.

We modeled both sudden and gradual extinctions. For sudden extinctions, P_o is constant until the extinction interval, when it drops instantaneously to zero. Gradual extinctions were characterized by probabilities of occurrence that declined linearly from 1 (at year 1) to zero (at T_{ext}, "true" extinction time). We do not claim that population declines in nature are necessarily linear, but because our goal is to evaluate differences in model performance resulting from the nature of the extinction (gradual vs. sudden), the linear decline provides a useful comparison. An exponentially declining model may be more realistic for some populations, but it proved to be less useful for our purpose in some cases because it can leave too few sightings for meaningful estimation of confidence intervals. We did, however, undertake exploratory analyses using an exponentially declining model and constraining $T_{ext} = 120$ years to ensure enough sightings. In this case the results were qualitatively similar to those of the linearly declining model (see *Discussion*).

Under perfect sampling, the probability of sampling a species if it is present at a locality would be 1.0. In practice, however, sampling is never perfect and its intensity tends to vary over time, sometimes systematically. In order to explore such sampling effects, we added a layer of sampling to the simulations that was superimposed on the population dynamics (gradual vs. sudden extinction). We modeled sampling effects using a parameter representing the probability of sampling an existing population (P_s), which varied over time according to one of five patterns: (1) uniform, (2) down, (3) down–up, (4) up, and (5) up–down (Fig. 1). Although all of these scenarios constitute only broad abstractions of reality, they do represent some likely real-world situations. Uniform sampling (type 1) is ideal, but probably the least realistic scenario over long time periods, although it may occur in areas with long-term monitoring programs. Sampling types 2 and 5, like many real sighting records, are heterogeneous over time in terms of sampling intensity. Sampling types 3 and 4 mimic intensive sampling efforts aimed to confirm the real extinctions of some species. In all cases, the sampling process continued up to the year 200, well beyond the real extinction time (20–120 yr). This is needed for two of the methods (class 2 methods) that explicitly require estimates of the pre- and post-extinction sampling intensity. For these methods, confidence intervals can be calculated only after ensuring a large enough post-extinction sampling. The uniform sampling scenario held P_s constant across the sampling period from the first year until year 200. In sampling type 2, P_s is maximum at T_1 and then declines linearly from T_1 to T_{ext}, where it reaches a stable minimum of $P_s = 0.2$. In type 3, P_s also declines linearly from T_1 to T_{ext}, but then it increases abruptly at T_{ext+1} to the initial P_s. In sampling type 4, P_s is zero at T_1 and then it increases linearly reaching a maximum at T_{ext}, which is maintained until the year 200. Finally, in sampling type 5, P_s starts from zero and increases linearly from T_1 and reaches a maximum P_s at T_{ext} but then it collapses abruptly to a $P_s = 0.2$ at T_{ext+1}.

The total probability that a species is sighted during a particular time interval is equal to the probability that it occurred, multiplied by the probability that it was sampled ($P_t = P_o \cdot P_s$). Although this total probability is all that matters for determining the record of sightings, we chose to model its two components separately so that we could tease apart the influences of actual occurrence and sampling effort on each estimation method. We ran simulations assuming relatively high sighting probabilities ($P_t = 0.8$, with $P_o = P_s = 0.894$) and relatively low sighting probabilities ($P_t = 0.2$, with $P_o = P_s = 0.447$); these values were chosen based on an empirical distribution of sighting rates (see *Discussion*). We also explored a broader range of values for these parameters, but similar results were obtained and are not shown here. With two levels of P_t, two kinds of extinctions (gradual and sudden), and five sampling scenarios, we report the results from a total of 20 different simulation scenarios.

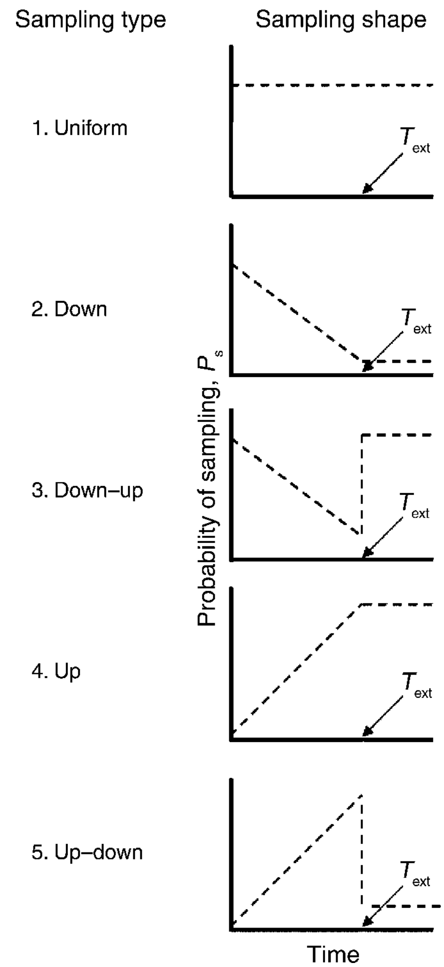

FIG. 1. Graphical descriptions of the five sampling types used in the simulations (for details, see *Methods: Simulated data set*). T_{ext} is the true time to extinction; probability of sampling is P_s. Uniform sampling (type 1) is ideal but probably the least realistic over long time periods. Types 2 and 5, like many real sighting records, are heterogeneous in sampling intensity over time. Types 3 and 4 mimic intensive sampling efforts aimed to confirm true extinctions.

Analyses

For each simulation run we estimated the upper 95% bound of the CI for extinction time using each of the seven methods listed in Table 1. Three of these methods (Solow 1993*a*, McCarthy 1998, McInerny et al. 2006) were originally formulated as extinction probabilities, and they were converted to CIs by rearranging the equations and solving for the time at which the probability of extinction equaled 0.05. The seven methods can be grouped into three major categories (classes 1, 2, and 3) based on the kind of assumptions that they make about sampling intensity over time. Class 1 methods make the strictest and most unrealistic assumption that sightings are a Poisson stationary process (Strauss and Sadler 1989, Solow 1993*a*, McInerny et al. 2006), with constant probabilities of occurrence and sampling through time, and an abrupt collapse to extinction. Class 2 methods (Marshall 1997, McCarthy 1998) relax the assumption of uniform probabilities of occurrence by explicitly accounting for temporal variation in the probability of sampling (the "recovery potential" of Marshall 1997). This is achieved by incorporating into the computations a measure (or proxy) for sampling intensity. In our analyses, we used the probability of sampling (P_s) as a proxy for sampling effort or recovery potential. While these methods would perform best if the total probability of sighting (P_t) were

TABLE 1. Methods used to estimate the upper bound (95th percentile) of the confidence interval of extinction times.

Method, by class	Source	Formula
Class 1		
S&S	Strauss and Sadler (1989)	$T_{ci} = T_n + \lambda R,\ \lambda = (1-\alpha)^{-1/(H-1)} - 1$
SOL	Solow (1993*a*)	$T_{ci} = \frac{T_n}{(1-\alpha)^{(1/H)}}$
MCY	McInerny et al. (2006)	$T_{ci} = T_n + \log_{[1-(H/T_n)]}(1-\alpha)$
Class 2		
MAR	Marshall (1997)	$\sum_{i=T_n}^{i=T_{ci}} e_i = \lambda \sum_{i=1}^{i=T_n} e_i$
MCC	McCarthy (1998)	$\sum_{i=1}^{i=T_{ci}} e_i = \frac{\sum_{i=1}^{i=T_n} e_i}{\alpha^{(1/H)}}$
Class 3		
S&R	Solow and Roberts (2003)	$T_{ci} = T_n + \left(\frac{1-\alpha}{\alpha}\right)(T_n - T_{n-1})$
R&S	Roberts and Solow (2003)	$T_{ci} = T_n + \frac{T_n - T_{n-H+1}}{c(\alpha) - 1}$
		$c(\alpha) = \left[\frac{-\log(\alpha/2)}{H}\right]^{-\hat{\nu}}$
		$\hat{\nu} = \frac{1}{H-1}\sum_{i=1}^{H-2}\log\frac{T_n - T_{n-H+1}}{T_n - T_{i+1}}$

Notes: Methods are grouped in three classes based on the assumptions made (see *Methods: Analyses*). We modified the original notations of the formulae in some cases so that the methods can be directly compared. In its original formulation, the MAR method is expressed as integrals, but given the discrete nature of most sighting records, summation may be more appropriate and is used here. For both class 2 methods, extinction time corresponds to the time at which the sampling level at the left side of the equations equals the right side of the equation. T_{ci} is the upper bound of the confidence interval of the extinction time, T_n is the time of the latest sighting, R is the total temporal span of the sightings, α is the confidence level (0.05), H is the total number of sightings, and e_i is the sampling effort (probability of sampling) in the ith year.

used, this parameter is generally not available for empirical sighting records because it depends on species abundance and persistence as well as sampling effort. In contrast, sampling effort can be estimated in a variety of ways (McCarthy 1998, Ungricht et al. 2005, van der Ree and McCarthy 2005, Farnsworth and Ogurcak 2006) and therefore its use as a proxy constitutes a more realistic test of these methods. Finally, class 3 methods (Roberts and Solow 2003, Solow and Roberts 2003) are the least restrictive in that they do not make any distributional assumptions, although the independence of sightings is still required. This informal taxonomy of methods is validated by our results, which show similar performance for methods within the same class. An Excel spreadsheet to compute the CIs for all the seven studied methods is provided (Supplement).

All of the methods evaluated here assume stable populations with a sudden extinction. Solow (1993*b*) developed a method to account for declining populations, but in that case the numerical solution of the upper bound of the confidence interval is not always possible (Solow 2005). Another approach is to use Bayesian methods to estimate extinction times (Ferraz et al. 2003), but their application requires a priori information about population dynamics that usually is not available for species where extinction times are estimated from sighting records. Thus these two methods were not included in our analyses.

Simulations were performed in an Excel spreadsheet using the PopTools module (Hood 2005). We created 40 000 runs for each of the 20 scenarios explored, excluding runs with fewer than five sightings, which is advisable for some of the methods (Strauss and Sadler 1989). Because the R&S method (Roberts and Solow 2003) assumes a Weibull extreme value distribution, analyzing sighting records with many observations may violate the asymptotic result upon which this method is based (Solow 2005). Our preliminary analyses showed that indeed the inclusion of a large number of sightings increased the upper bounds of the estimates. Although the optimal number of included sightings is still an open question (Solow 2005), we followed a previous study (Roberts and Solow 2003) in using, at most, the 10 latest sightings with this method.

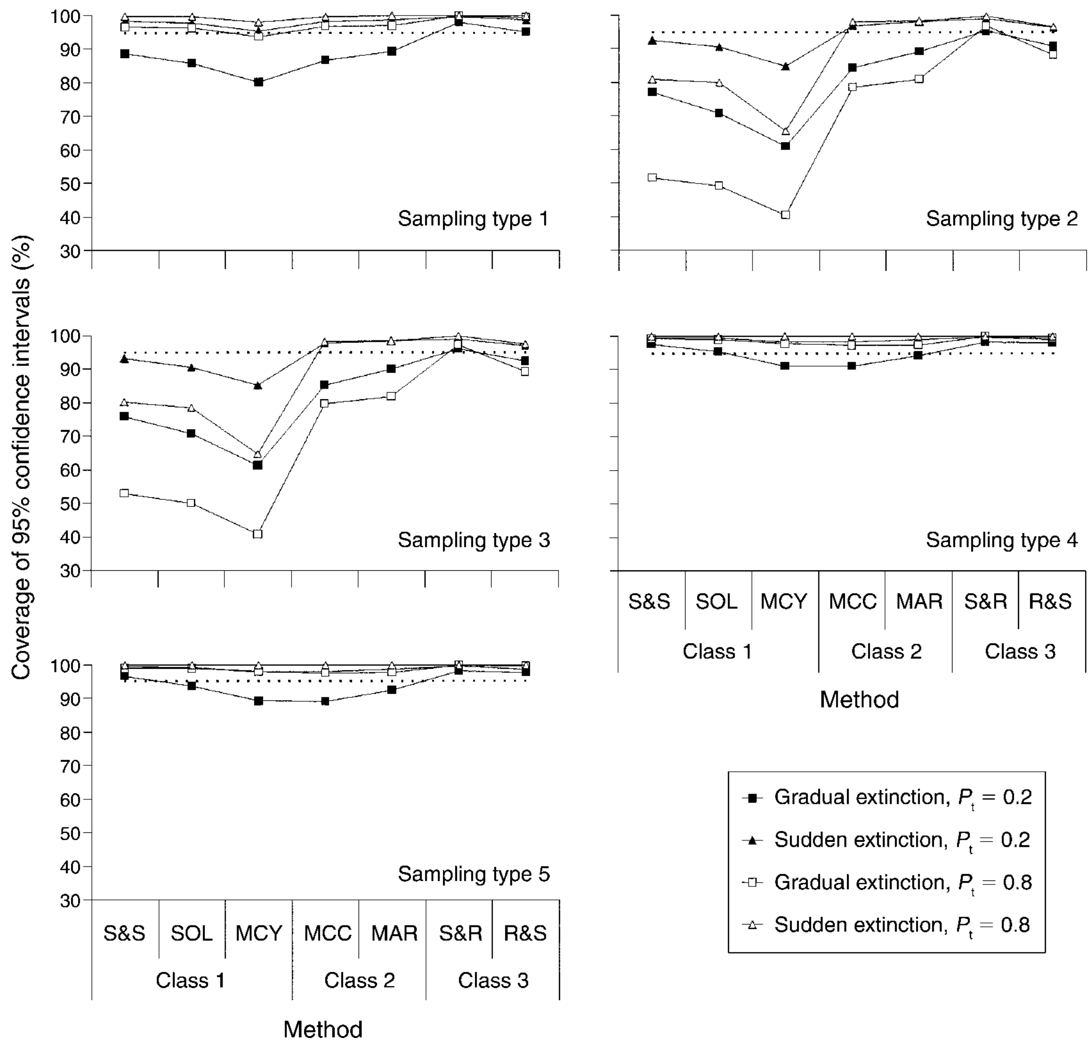

FIG. 2. Coverage for 95% confidence intervals of the upper bounds of extinction times, according seven methods under different simulation scenarios: gradual vs. sudden extinction, and total detection probability (P_t) of 0.2 vs. 0.8 (see Fig. 1 and *Methods: Simulated data set*). The broken line shows the nominal 95% indicating a perfect coverage. The methods are: S&S (Strauss and Sadler 1989), SOL (Solow 1993*a*), MCY (McInerny et al. 2006), S&R (Solow and Roberts 2003), R&S (Roberts and Solow 2003), MAR (Marshall 1997), and MCC (McCarthy 1998). These methods are grouped based on assumptions about sampling intensity. Class 1 methods make the strictest, most unrealistic assumption that sightings are a Poisson stationary process, with constant probabilities of occurrence and sampling through time, and an abrupt collapse to extinction. Class 2 methods relax the assumption of uniform probabilities of occurrence by accounting for temporal variation in the probability of sampling. Class 3 methods are the least restrictive; they make no distributional assumptions, although they still require independence of sightings.

The performance of each method under all of the different scenarios was evaluated according to the coverage of the resulting confidence intervals. Coverage is defined as the probability that the true parameter value occurs within the last sighting and the upper bound of the 95% CI. In the case of our simulations, the coverage of a method is equal to the percentage of runs in which the true extinction time occurs within this interval (i.e., the upper 95% confidence bound is equal to or higher than the true extinction time, $T_{ci} \geq T_{ext}$). Ideally, coverage should coincide with the nominal level of the confidence interval, so that, on average, 95% of simulation runs should have true extinction times that fall between the last sighting and the upper 95% CI. If a method produces confidence limits that are systematically too broad, coverage will be greater than the nominal percentage, and the procedure is thus too conservative, inflating Type II error. On the other hand, if a method tends to produce overly narrow confidence intervals, its coverage will be below the nominal percentage, increasing the incidence of Type I errors. In order to evaluate the consistency of each method, we ranked them according to the coverage under each scenario, and for each method we calculated Spearman rank correlations among all possible pairs of scenarios. Finally, we analyzed the effect of the number of

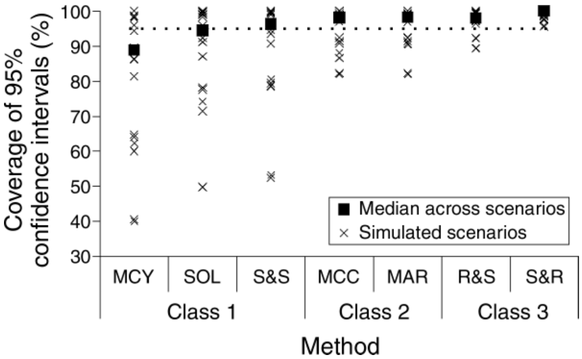

FIG. 3. Median coverage of 95% CI for the three classes of methods, calculated across the 20 simulated scenarios (see Fig. 2). The broken line shows the nominal 95% indicating a perfect coverage. See Fig. 2 and Table 1 for method abbreviations.

sightings (above and below the median of 14 sightings in the simulations), and the sighting rate (above and below the median sighting rate in the simulations = 0.33 sightings/year) on the coverage of one method representing each family of methods. CI values were calculated under two contrasting scenarios: best case (sampling type 1, sudden extinction, $P_t = 0.8$) and worst case (sampling type 2, gradual extinction, $P_t = 0.8$).

RESULTS

Realized coverage of 95% confidence intervals varied widely under different simulation scenarios and methods and the results are summarized in Fig. 2. In the simplest case when the assumptions of all of the methods are met (uniform sampling effort [type 1] and sudden extinction), all methods produced confidence intervals that were overly broad (>95% coverage). When probabilities of sightings were high ($P_t = 0.8$), most methods yielded conservatively broad confidence intervals, with coverages close to 100%. Coverages were lower with gradual extinction and lowered sighting probabilities ($P_t = 0.2$), such that the combination of both of these factors caused all class 1 and class 2 methods to produce confidence intervals that were too narrow (<95% coverage).

Sampling scenarios with probabilities of sampling that increased toward the real extinction time (types 4 and 5) produced results similar to those produced under constant sampling (type 1), with confidence intervals that were rather broad under most simulated circumstances (Fig. 2). However, when sampling efforts decreased over time (types 2 and 3), method performance was often drastically altered. For these sampling scenarios, realized coverage of the class 1 methods plummeted below the nominal 95% level, sometimes markedly so. This effect was more pronounced when extinctions were gradual and sighting probabilities (P_t) were high. Class 2 and class 3 methods were less affected by decreasing sampling intensities (types 2 and 3), but coverages still decreased, on average, and in some cases dipped below the nominal 95% level (Fig. 2).

Qualitatively, these patterns can be attributed to a few effects. Under normal circumstances all of the methods produce confidence intervals that are somewhat broad. When the probability of observing a species decreases over time, species are less likely to be observed during the intervals immediately preceding extinction. As a result, methods tend to overstate the evidence for early extinction and therefore produce confidence intervals that are too narrow. The effect is similar, regardless of whether the decrease in sighting probabilities is caused by gradual extinction (decreasing P_o) or by declining sampling effort (P_s). Class 2 methods can correctly account for the latter but not the former (compare results for gradual vs. sudden extinction with decreasing sampling: type 2 in Fig. 2). That this effect is actually worse when sighting levels are higher ($P_t = 0.8$) is perhaps surprising, but explainable because more early

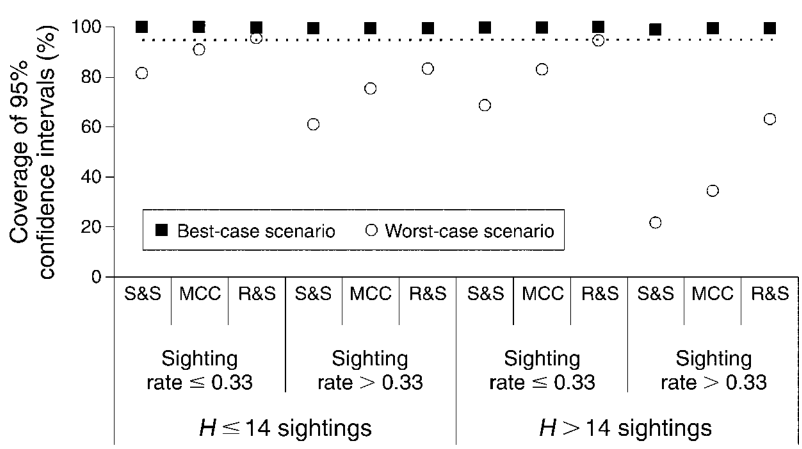

FIG. 4. Effect of the number of sightings (H) and sighting rate (number of sightings/yr) on the coverage of the 95% confidence intervals obtained for three methods representing primary (S&S), secondary (MCC), and tertiary (R&S) methods, under best-case scenarios (sampling type 1, sudden extinction, $P_t = 0.8$) and worst-case scenarios (sampling type 2, gradual extinction, $P_t = 0.8$). The cutoff level for H and sighting rate correspond to the respective median values. See Fig. 2 and Table 1 for abbreviations.

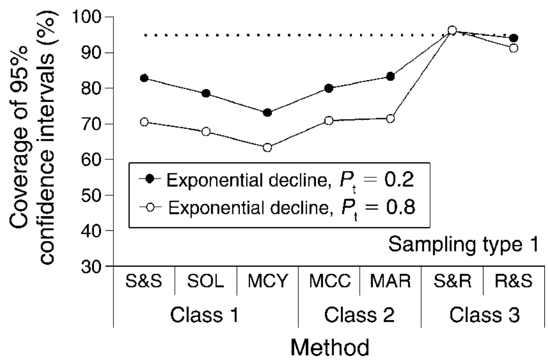

FIG. 5. Coverage for 95% confidence intervals of the upper bounds of extinction times, assuming an exponential decline and a type 1 (uniform) sampling. In order to get enough sightings, the real extinction time T_{ext} was constrained to be 120 years.

sightings (when positive sightings are more probable) seem to more strongly support erroneously early extinction estimates.

Performance was generally similar within each family of methods (Figs. 2 and 3). Median coverage across the 20 simulated scenarios increased from class 1 to class 3 methods, coupled with a decrease in the among-scenario variation in realized coverage. This ranking of methods was very consistent across scenarios, with strong Spearman rank correlations among all possible pairs of scenarios (median $r_S = 0.70$, 5th percentile $r_S = 0.29$, 95th percentile $r_S = 1.0$, $n = 190$). Class 1 methods, especially the MCY method, produced much narrower confidence limits compared to other methods. At the other extreme, class 3 methods nearly always yielded very broad confidence intervals (median coverage $> 98\%$).

Both sighting number and sighting rate had an effect on coverage, but the results varied according the case scenario and methods (Fig. 4). Under the worst-case scenario (sampling type 2, gradual extinction, $P_t = 0.80$), confidence limits were generally too liberal with many sightings ($H > 14$ sightings) and high sighting rates (>0.33); the S&S method was much more affected by changes in those variables than other methods. At high sighting number and high sighting rate, all methods performed poorly ($<63\%$ coverage), with coverage values as low as 21% in the S&S. Overall, the R&S method yielded more robust estimations under different combinations of sighting number and sighting rate. In the best-case scenario (sampling type 1, sudden extinction, $P_t = 0.8$), the number of the sightings had no apparent effect on the coverage, but the increase in the sighting rates yielded coverage values closer to the nominal 95% in all of the methods analyzed.

DISCUSSION

Our simulations demonstrate that the ability of existing methods to correctly estimate the timing of extinctions varies with the type of extinction (sudden vs. gradual) and the nature of sampling effort over time. These results are consistent with previous analyses that have partially addressed this issue (Strauss and Sadler 1989, Solow 1993*b*, Burgman et al. 1995, 2000, Grogan and Boreman 1998, McCarthy 1998, McInerny et al. 2006).

Within the range of conditions simulated here, the most important factor in determining the accuracy of confidence limits is whether there has been a historical decline in sampling intensity. When sampling efforts decline over time, there is a substantial chance of falsely concluding that a species is extinct, especially if class 1 methods are used. For example, if efforts devoted to build natural history collections that provide temporal data on species occurrences decline over time, as has been documented in some cases (McCarthy 1998, Burgman et al. 2000, Hedenas et al. 2002), the ability of these methods to accurately estimate past extinction

TABLE 2. Comparison of the coverage of 95% confidence intervals (CIs) of upper bounds of extinction times vs. the mean length of the upper bound of CIs (in years), for each method under the best-case and worst-case scenario.

Method, by class	Best-case scenario: Coverage (%)	Best-case scenario: Mean length of CIs (yr)	Worst-case scenario: Coverage (%)	Worst-case scenario: Mean length of CIs (yr)
Class 1				
S&S	97	9	52	19
SOL	96	8	49	15
MCY	94	6	40	11
Class 2				
MCC	97	9	79	31
MAR	97	9	81	31
Class 3				
S&R	100	47	97	151
R&S	100	15	88	67

Notes: The best-case scenario is sampling type 1, sudden extinction, and $P_t = 0.8$. The worst-case scenario is sampling type 2, gradual extinction, and $P_t = 0.8$. The nominal 95% indicates a perfect coverage.

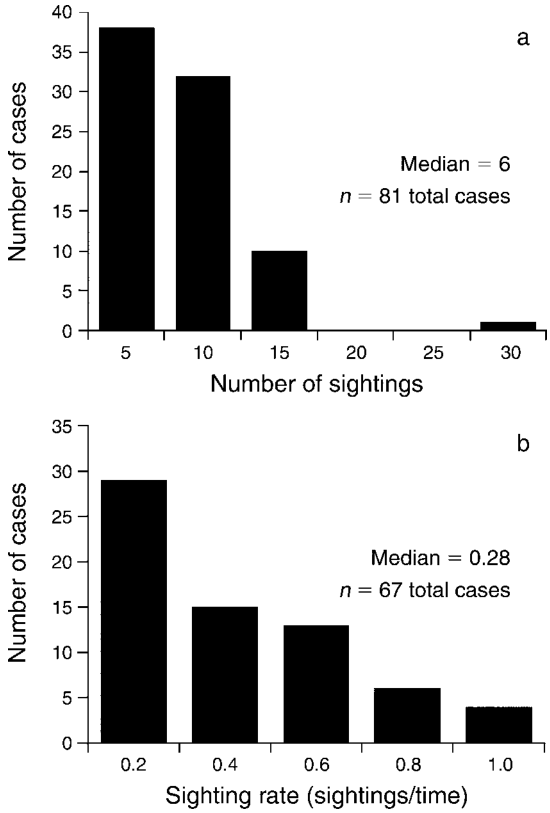

FIG. 6. Frequency distribution of (a) the number of sightings and (b) sighting rates used to evaluate the extinction time and probabilities reported in the literature. Cases reported in the literature include both endangered and non-endangered species. Data are from Solow (1993*b*, 2005), Burgman et al. (1995, 2000), Roberts and Solow (2003), Solow and Roberts (2003), McInerny et al. (2006), Roberts (2006), Roberts and Kitchener (2006), and Solow et al. (2006*a*).

times may be limited in many situations. This problem is exacerbated when populations gradually decline to extinction: as both sampling and occurrence probabilities decline, the sighting probability declines even faster (see also Burgman et al. 2000). The extreme situation of an exponential decline may potentially amplify the problem, although our exploratory analyses (Fig. 5) suggest that under a uniform sampling, and given enough sightings, the effect of the exponential decline may be comparable to a linear decline with $P_t = 0.2$. This is not surprising, because a both linear and exponential extinction process violates the main assumption of these methods that the sighting series should be produced by a random stationary process (Solow 1993*a*, 2005, Burgman et al. 1995).

Class 1 methods are generally outperformed by class 2 and class 3 methods that are far less sensitive to temporal variation in sampling intensity. On the other hand, although class 3 methods are robust to varying sampling rates and scenarios, they are also the most conservative, often producing confidence intervals with coverage much higher than the nominal value. This can be advantageous if one wishes to be cautious in declaring a species extinct because false inferences of extinction are potentially problematic (Roberts 2006, Roberts and Kitchener 2006, Solow et al. 2006*a*). On the other hand, these methods will also risk frequent Type II errors (an already extinct population is still considered as extant, e.g., Solow 2005). For instance, the mean length of CIs obtained for class 3 methods (especially for S&R method) can be more than 10-fold larger than for class 1 and class 2 methods (Table 2). Along this continuum, class 2 methods seem to provide a better balance between the two types of error, especially when a locality or region has been well sampled after the extinction of the target species. However, when the interval after the true extinction is poorly sampled, the length of confidence intervals under class 2 methods increases enormously, and in extreme cases they may not even be estimated (Labandeira et al. 2002). Moreover, the choice of the best measure of sampling effort or "recovery potential" (Marshall 1997, McCarthy 1998, Holland 2003) is not always clear. Several proxies have been proposed, including the total number of species and the total occurrences recorded, with applications to both ecological (McCarthy 1998, Ungricht et al. 2005, van der Ree and McCarthy 2005, Farnsworth and Ogurcak 2006) and paleoecological data sets (Labandeira et al. 2002, Holland 2003). In our analyses, we used the simulated probability of sampling as a measure of recovery potential; a separate issue that we have not explored here is the performance of class 2 methods when the metric of recovery potential is noisy or misleading.

The quality of the data sets used for the estimation of extinction time, in terms of the length (number of sightings) and rate of the sighting record, clearly has an effect on our ability to estimate the timing of extinctions, although effects were modulated by the sampling type and type of extinction. More specifically, our simulations suggest that sighting rate, a measure that integrates ecological and sampling factors affecting occupancy and probabilities of sampling, may affect the reliability of the estimations, confirming previous analyses (Burgman et al. 2000, Solow and Roberts 2003, McInerny et al. 2006). High sighting rates (e.g., >0.33 sightings/year in our simulations) are expected in temporally persistent and/or well-sampled populations, so relatively small temporal gaps may be interpreted as true extinctions. This characteristic will improve estimations under an ideal best-case scenario (e.g., sudden extinction, sampling type 1), but under a worst-case scenario (e.g., gradual extinction, sampling type 2), this same characteristic will increase the chances of falsely inferring extinctions, especially in longer time series ($H > 14$ sightings); see Fig. 4. Very short time series may also lead to misleading results, e.g., when using fewer than six sightings (Strauss and Sadler 1989), and a search of the literature shows that the frequency distribution of the number of sightings used to evaluate extinction probabilities tends to be strongly right skewed and a number of empirical studies are based on fewer than six sightings (Fig. 6a). In those cases, the upper CI could be seriously inflated (Strauss and Sadler 1989). In addition, the literature search reveals that a large percentage of empirical studies of extinctions (>50% of cases) were based on data sets with relatively high sighting rates (i.e., >0.28 sightings/time; Fig. 6b), which may have had an important effect on the estimations if the sampling efforts were not uniform across the study intervals.

Because of the nature of historical sighting records, improving the number of sightings or the sighting rate is not an option in most cases. Similarly, detailed knowledge of population dynamics (i.e., probabilities of occupancy), sampling effort (i.e., probabilities of sampling), and temporal trends in these variables is also not readily available for many species. Yet there is a pressing need to better estimate the timing of extinctions of species and populations; our results provide some concrete recommendations regarding the choice of methods, given the nature of the available data. Using a battery of different methods to estimate confidence limits may be a reasonable way to guard against the shortcomings of any individual method (Burgman et al. 2000, Ferraz et al. 2003, Ungricht et al. 2005, van der Ree and McCarthy 2005, Farnsworth and Ogurcak 2006, Robbirt et al. 2006, Roberts and Kitchener 2006), but our results show that there are real differences in performance among methods and that, under some circumstances, some methods are positively misleading. Thus using all available methods for any given data set may not always be advisable. Instead our simulations suggest that the most important factor to consider in choosing a method is whether sampling efforts have declined over time or not. If sampling has declined, class 3 methods seem to give the best results overall and class 2 methods may be acceptable if extinction was sudden. If sampling has not declined over time, class 3 methods generally appear to be overly conservative and will often fail to detect true extinctions. Under such circumstances, class 2 methods may be better. Class 1 methods, on the other hand, should only be used if sampling has not decreased over time and sampling rates are high (e.g., for species where long-term biological monitoring programs are in place). Even then, class 2 methods appear to be better suited unless a reasonable proxy for sampling intensity is lacking. Finally, if the emphasis is to avoid false extinctions, then class 3 methods may be preferred, but this choice comes with much reduced power to detect true extinction events.

In summary, reliably estimating extinction timing from sighting data remains a difficult problem, given the nature of the extinction process and incomplete information about how it unfolds. However, many of the existing methods can provide reliable estimates of extinction times as long as the choice of the method is determined by the nature of the underlying sighting data.

Acknowledgments

We thank two anonymous reviewers for insightful comments that substantially improved the manuscript. This work was supported by a grant from NOAA-California Sea Grant to K. Roy, and by a FONDECYT grant (11070147) to M. M. Rivadeneira.

Literature Cited

Burgman, M. A., R. C. Grimson, and S. Ferson. 1995. Inferring threat from scientific collections. Conservation Biology 9: 923–928.

Burgman, M. A., B. R. Maslin, D. Andrewartha, M. R. Keatley, C. Boek, and M. McCarthy. 2000. Inferring threat from scientific collections: power tests and an application to Western Australian *Acacia* species. Pages 7–26 *in* S. Ferson and M. A. Burgman, editors. Quantitative methods for conservation biology. Springer-Verlag, New York, New York, USA.

Connolly, S. R., and A. I. Miller. 2001. Joint estimation of sampling and turnover rates from fossil databases: capture–mark–recapture methods revisited. Paleobiology 27:751–767.

Dorazio, R. M. 2007. On the choice of statistical models for estimating occurrence and extinction from animal surveys. Ecology 88:2773–2782.

Dulvy, N. K., J. R. Ellis, N. B. Goodwin, A. Grant, J. D. Reynolds, and S. Jennings. 2004. Methods of assessing extinction risk in marine fishes. Fish and Fisheries 5:255–276.

Farnsworth, E. J., and D. E. Ogurcak. 2006. Biogeography and decline of rare plants in New England: historical evidence and contemporary monitoring. Ecological Applications 16: 1327–1337.

Ferraz, G., G. J. Russell, P. C. Stouffer, R. O. Bierregaard, S. L. Pimm, and T. E. Lovejoy. 2003. Rates of species loss from Amazonian forest fragments. Proceedings of the National Academy of Sciences (USA) 100:14069–14073.

Grogan, C. S., and J. Boreman. 1998. Estimating the probability that historical populations of fish species are extirpated. North American Journal of Fisheries Management 18:522–529.

Hayek, L. C., and E. Bura. 2001. On the ends of the taxon range problem. Pages 221–244 *in* J. B. C. Jackson, S. Lidgard, and F. K. McKinney, editors. Evolutionary patterns: growth, form, and tempo in the fossil record. University of Chicago Press, Chicago, Illinois, USA.

Hedenas, L., I. Bisang, A. Tehler, M. Hamnede, K. Jaederfelt, and G. Odelvik. 2002. A herbarium-based method for estimates of temporal frequency changes: mosses in Sweden. Biological Conservation 105:321–331.

Holland, S. M. 2003. Confidence limits on fossil ranges that account for facies changes. Paleobiology 29:468–479.

Hood, G. M. 2005. PopTools version 2.6.7. ⟨http://www.cse.csiro.au/poptools⟩

Jablonski, D. 1986. Causes and consequences of mass extinctions. Pages 183–229 *in* D. K. Elliott, editor. Dynamics of extinction. John Wiley, New York, New York, USA.

Labandeira, C. C., K. R. Johnson, and P. Wilf. 2002. Impact of the terminal Cretaceous event on plant–insect associations. Proceedings of the National Academy of Sciences (USA) 99: 2061–2066.

MacKenzie, D. I., J. D. Nichols, J. E. Hines, M. G. Knutson, and A. B. Franklin. 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. Ecology 84:2200–2207.

MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. H. Pollock, L. L. Bailey, and J. E. Hines. 2006. Occupancy estimation

and modeling. Inferring patterns and dynamics of species occurrence. Elsevier, Amsterdam, The Netherlands.
Marshall, C. R. 1990. Confidence intervals on stratigraphic ranges. Paleobiology 16:1–10.
Marshall, C. R. 1994. Confidence intervals on stratigraphic ranges: partial relaxation of the assumption of randomly distributed fossil horizons. Paleobiology 20:459–469.
Marshall, C. R. 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 of western European Tethys. Science 274:1360–1363.
McCarthy, M. A. 1998. Identifying declining and threatened species with museum data. Biological Conservation 83:9–17.
McInerny, G. J., D. L. Roberts, A. J. Davy, and P. J. Cribb. 2006. Significance of sighting rate in inferring extinction and threat. Conservation Biology 20:562–567.
Nichols, J. D., T. Boulinier, J. E. Hines, K. H. Pollock, and J. R. Sauer. 1998. Estimating rates of local species extinction, colonization, and turnover in animal communities. Ecological Applications 8:1213–1225.
Nichols, J. D., and K. H. Pollock. 1983. Estimating taxonomic diversity, extinction rates, and speciation rates from fossil data using capture–recapture models. Paleobiology 9:150–163.
Robbirt, K. M., D. L. Roberts, and J. A. Hawkins. 2006. Comparing IUCN and probabilistic assessments of threat: Do IUCN red list criteria conflate rarity and threat? Biodiversity and Conservation 15:1903–1912.
Roberts, D. L. 2006. Extinct or possibly extinct? Science 312: 997–998.
Roberts, D. L., and A. C. Kitchener. 2006. Inferring extinction from biological records: Were we too quick to write off Miss Waldron's Red Colobus Monkey (*Piliocolobus badius waldronae*)? Biological Conservation 128:285–287.
Roberts, D. L., and A. R. Solow. 2003. Flightless birds: When did the dodo become extinct? Nature 426:245–245.
Shaffer, H. B., R. N. Fisher, and C. Davidson. 1998. The role of natural history collections in documenting species declines. Trends in Ecology and Evolution 13:27–30.
Solow, A. R. 1993*a*. Inferring extinction from sighting data. Ecology 74:962–964.
Solow, A. R. 1993*b*. Inferring extinction in a declining population. Journal of Mathematical Biology 32:79–82.
Solow, A. R. 2005. Inferring extinction from a sighting record. Mathematical Biosciences 195:47–55.
Solow, A. R., A. C. Kitchener, D. L. Roberts, and J. D. S. Birks. 2006*a*. Rediscovery of the Scottish polecat, *Mustela putorius*: Survival or reintroduction? Biological Conservation 128:574–575.
Solow, A. R., and D. L. Roberts. 2003. A nonparametric test for extinction based on a sighting record. Ecology 84:1329–1332.
Solow, A. R., D. L. Roberts, and K. M. Robbirt. 2006*b*. On the Pleistocene extinctions of Alaskan mammoths and horses. Proceedings of the National Academy of Sciences (USA) 103: 7351–7353.
Strauss, D., and P. M. Sadler. 1989. Classical confidence intervals and Bayesian probability estimates for ends of local taxon ranges. Mathematical Geology 21:411–421.
Ungricht, S., J. Y. Rasplus, and F. Kjellberg. 2005. Extinction threat evaluation of endemic fig trees of New Caledonia: priority assessment for taxonomy and conservation with herbarium collections. Biodiversity and Conservation 14: 205–232.
van der Ree, R., and M. A. McCarthy. 2005. Inferring persistence of indigenous mammals in response to urbanisation. Animal Conservation 8:309–319.
Wang, S. C., and C. R. Marshall. 2004. Improved confidence intervals for estimating the position of a mass extinction boundary. Paleobiology 30:5–18.

SUPPLEMENT

A spreadsheet to estimate extinction times based on a sighting record (*Ecological Archives* E090-084-S1).