Title: *Recent environmental change may affect accurate inference of extinction*

Running head: *Accurately inferring extinction status*

Word count: 5,994

Keywords: species loss, method comparison, environmental change, wild populations, microcosm

Authors:

Christopher F Clements*

Department of Animal and Plant Sciences, University of Sheffield, Sheffield, S10 2TN, United Kingdom.

Email: *c.clements@shef.ac.uk,* Tel: *(+44)7708945975*

Ben Collen

Centre for Biodiversity & Environment Research, University College London, Gower Street, London WC1E 6BT, United Kingdom

Tim M. Blackburn

Institute of Zoology, Zoological Society of London, Regent's Park, London NW1 4RY, United Kingdom and Distinguished Scientist Fellowship Program, King Saud University, P.O. Box 2455, Riyadh 1145, Saudi Arabia

Owen L. Petchey

Institute of Evolutionary Biology and Environmental Studies, The University of Zurich, Zurich, CH-8057, Switzerland

*corresponding author

Abstract

Correctly classifying a species as extinct or extant is of critical importance if current rates of biodiversity loss are to be accurately quantified. Observing an extinction event is rare, so in many cases extinction status is inferred using methods based on the analysis of records of historic sighting events. The accuracy of such methods is difficult to test. However, recent experiments using microcosm communities suggests that the rate at which a population declines to extinction, potentially driven by varying environmental conditions, may alter our ability accurately to infer extinction status. We tested how the rate of population decline, driven by historic environmental change, alters the accuracy of six commonly used sighting based methods for inferring extinction, using data from small-scale experimental communities and recorded wild population extirpations. We assessed how accuracy of the different methods depends on rate of population decline, search efforts, and number of sighting events recorded. Although the rate of populations decline affected the accuracy of inferred extinction dates, so did the historic population size of the species; faster declines produced more accurate inferred dates of extinction, but only when population sizes were higher. Optimal Linear Estimation (OLE) offered the most reliable and robust estimates, though no single method performed best in all situations, and it may be appropriate to use a different method if information regarding historic search efforts is available. Importantly, we show that OLE provided the most accurate estimates of extinction when the number of sighting events used was >10, and future use of this method should take this into account. Data from experimental populations provide added insight into testing techniques to discern wild extirpation events. Care should be taken designing such experiments to more closely mirror the abundance dynamics of populations that suffer real world extirpation events.

Introduction

Reducing global biodiversity loss in the face of unprecedented population extirpation and species extinction has become a fundamental goal for conservation groups. However, whilst current extinction rates are thought to be much higher than those recorded in the fossil record (Barnosky *et al.* 2011), quantifying the exact rate of species loss, despite much invested effort, remains problematic (Fisher & Blomberg 2011; Clements et al. 2013). This is, in part, due to the difficulty of observing extinction, i.e., the absence of something that is otherwise rarely seen, and this difficulty has given rise to many techniques that attempt to allow historic extinction events to be inferred, rather than observed directly (Solow 1993b, 2005; Burgman et al. 1995; McCarthy 1998; Roberts & Solow 2003; Gotelli et al. 2011). Given the often limited information available on many species, such methods have often concentrated on inferring extinction based on historic sighting events data (e.g. Solow 1993a, 2005; Roberts & Solow 2003; Solow & Roberts 2003; McPherson & Myers 2009). Recent work has suggested that such quantitative methods could be used to inform decisions on whether to classify species as extinct (Collen *et al.* 2010), however the accuracy of these methods remains difficult to test. Traditionally such tests have been tackled with either data from wild populations that may have suffered local extinction events (e.g. Collen *et al.* 2010), or with data from simulated populations (e.g. Rivadeneira *et al.* 2009). Recently, we have used experimental microcosm communities to provide detailed abundance time series data for species where the date of extinction can be accurately observed (Clements *et al.* 2013). Such an approach allows one to test the accuracy of estimates because the actual date of extinction is precisely known, something that is rarely possible with wild populations.

The rate at which a population declines to extinction can play an important role in determining how accurately a method for inferring extinction performs (Rivadeneira et al. 2009; Collen et al. 2010; Clements et al. 2013). Where species persist at low density for a lengthy period of time (and thus are rarely observed), estimates are worse than when the species falls rapidly to extinction. Thus, historic pressures on a species (be those abiotic, such as temperature change or habitat loss, or biotic, such as an invasive species or disease) that increase the rate at which a species declines may alter our ability to judge accurately whether the species has in fact been lost, and what time frame that may have occurred over.

Data that have been collected on wild populations have shown that both the identity of the species (and thus life history) as well as the nature of the threat can alter the rate of population decline (Weimerskirch & Jouventin 1987; Laurance et al. 1996; Di Fonzo et al. 2013). For example, Laurence *et al.* (1996) showed distinct differences in the rates of rapid disease driven population declines of four species of rain forest dwelling frogs, whilst Weimerskirch & Jouventin (1987) recorded differences in the rates of population decline of *Diomedea exulans*, the Wandering Albatross, between islands, probably as a result of each island’s location in relation to fishing areas. Given the high rates of environmental change over the last 100 years (Crowley 2000) the potential for factors that govern the rates of a species’ decline to alter our ability to infer whether a species is extinct is of concern, and quantifying this effect is an issue that is may affect our current understanding of the scale of biodiversity loss.

Here we utilise small-scale experimental communities to test whether there is a negative correlation between rate of population decline and the magnitude of the error of inferred extinction dates produced by six commonly applied methods. We then applied the same techniques to eight historic wild population extirpations, to see whether the results produced when using the microcosm data may also apply to real-world data. The experimental populations experienced various rates of environmental change, which altered their rate of their decline and time of extinction. The wild population data were selected to include a variety of rates of population decline driven by a number of different processes (including habitat loss, disease and extreme weather events). Given time series of the abundance of these experimental and wild populations we then generated time series of sighting events using three search regimes (Rivadeneira et al. 2009; Clements et al. 2013), and examined the effect of the rate of population decline, as well as the effect of search regime and the number of sighting events the calculations used, on the accuracy of inferred dates of extinction, and compare the robustness of the six techniques. Finally, we make recommendations about the use of such techniques in real-world scenarios, based on the results presented here.

Methods

Experimental set up and sampling

Microcosms were lidded petri dishes (Ø 100mm, height 25mm) containing 50ml of medium. The medium consisted of 0.05g/L of crushed protist pellets (Carolina Biological Supply, Burlington, NC), providing organic nutrients, and Chalkley's solution (Thompson *et al.* 1988), containing essential salts. On day −14 the medium was inoculated with the bacteria *Bacillus cereus* and *Serratia marcescens* and incubated at 20^{o}C. On day -10, a volume of high-density stock culture containing ~200 individuals of the ciliate bacterivore *Loxocephalus sp.* was added to each litre of medium. This culture was sampled every two days, and the experiment started (day 0) when the density of *Loxocephalus* had reached approximately the carrying capacity of the medium (i.e. exponential growth had stopped). On day zero the medium was homogenised, and 50ml added to each of 27 petri dishes (three replicate populations of nine temperature treatments).

The temperature treatments comprised: one treatment kept at a constant 20^{o}C for the duration of the experiment (treatment *C*), four treatments that were heated at different rates, and four treatments that were cooled at different rates (Fig. 1a). Cooling and heating were achieved by moving replicate populations between nine incubators set at 1.5^{o}C increments from 14^{o}C to 26^{o}C. The four heated treatments were: 1) increasing 0.5^{o}C/week (*I0.5*), 2) increasing 0.75^{o}C/week (*I0.75*), 3) increasing 1.5^{o}C/week (*I1.5*) and 4) increasing 3^{o}C/week (*I3*). The four treatments that decreased in temperature (*D*) mirrored the heated, and were thus *D0.5*, *D0.75*, *D1.5* and *D3*.

Microcosms were sampled to estimate population abundances twice per week for 10 weeks. Sampling was based on the protocol of Lawler & Morin (1993); microcosms were homogenised by repeat pipetting of the medium, and then a known volume (0.1-0.3ml) extracted using a an adjustable-volume pipette. The individuals within this known volume were then counted under a stereoscopic microscope (7.5-30x magnification), and the total population in the microcosm estimated. When densities became very low the whole microcosm was placed under the microscope and the individuals counted. A species was recorded as extinct if, on two consecutive sampling days, no individuals were observed after 5 minutes of searching. Previously this method has been shown to reliably identify when a species has gone extinct (Clements et al. 2013), and no populations that were initially recorded as extinct were re-observed at the next sampling occasion. All medium was replaced after counting, and any evaporative loss (checked with a balance) was replaced with distilled water.

Wild population data

Data on population dynamics followed by extirpation events were collected from three sources: a literature search (using Google Scholar with search terms such as “extirpation”, “population extinction” and “extinction dynamics”), the Living Planet Database (Collen et al. 2009), and Fagan and Holmes (2006). From these datasets we selected eight time series (Parr 1992; Burrows et al. 1995; Laurance et al. 1996; Fagan & Holmes 2006): one mammal (*Lycaon pictus*), four birds (*Corvus hawaiiensis, Crex crex, Grus americana, Pluvialis apricaria*), and three amphibians (*Litoria nannotis, Litoria rheocola, Taudactylus acutirostris*), each with at least seven recorded population abundances prior to a recorded extirpation event (a recorded population count of 0). These time series covered a range of rates of population decline from slow to fast (estimated by fitting linear regressions to the abundance data, Fig. 2), hypothesised to be caused by a variety of factors including extreme weather events, disease, habitat loss and degradation, and invasive species. These rates of decline ranged from the very rapid (e.g. *Litoria rheocola*, the Common Mist Frog) where approximately 40% of the initial population was lost per year, to the relatively slow (e.g. *Crex crex*, the Corncrake), where the population declined by roughly 16% of the initial population per year.

Creating sighting events

Abundance data from replicates of *Loxocephalus* where extinction was observed (all populations except those in the treatments *D1.5* and *D3*, where no extinctions were recorded), and wild populations, were converted into sighting data based on the method proposed by Clements et al. (2013) (see below). To these records of abundance through time three simulated search regimes were applied: (i) constant, (ii) increasing and (iii) decreasing effort. The "constant" search regime was simulated with search efforts (the fraction of the habitat search) of 0.01 to 0.95, in 0.01 steps, held constant through time. The "increasing" search regime had a randomly assigned initial search effort, and then increased by a random fraction at each time step, until the search effort reached 0.95 after which it remained constant. "Decreasing" mirrored the "increasing" search effort, but the fraction of the habitat decreased through time. These simulated search efforts at each point in time were then used to generate series of sighting events. Multiplying the search effort (the fraction of habitat searched) by the total number of individuals in the entire habitat gave the expected number of individuals observed. The actual number observed was drawn from a Poisson distribution with mean set to this expectation.

As in Clements et al. (2013), these sampling regimes produced regular sighting events when search effort or abundance was high. However in reality this is probably unrealistic, as sampling of wildlife populations is often sporadic (Turvey *et al.* 2007). Thus, two search "regularities" were simulated, "regular" sampling (as above), and "irregular" sampling. Irregular sampling was implemented identically to the first, but with every time point where observations occurred having a 50% probability of being used. This was done for both the experimental and wild population data. All analyses present results that include data from both regular and irregular sampling, with the results of the effect of regularity of sampling on the accuracy of extinction estimates presented in Appendix S1.

Hereafter we refer to the times at which sightings were recorded as "sighting events", these are days in the experimental system, and months or years in the wild population data. At each of these points in time there are a number of observations generated, which depend on the abundance of the population and the amount of the habitat searched; these are referred to as "sightings". If there were less than four sighting events then estimates were not made. All methods for inferring extinction were tested with identical data each time a set of sighting events was produced.

The sightings records derived above were used to test the six sighting-based methods for inferring historical extinction currently included in the R package "sExtinct" (Appendix S2). These methods do not explicitly take into account the search effort that generated a sighting events (although methods that do this do exist, e.g. Marshall (1997) & McCarthy (1998)). Thus, given that these methods will be applied where search efforts are inherently unknown, it is especially important to gauge their performance under various search efforts and drivers of predictive error. These will be referred to by simplifications of the function names in the R package, and are as follows: (i) Burgman (Burgman *et al.* 1995), (ii) OLE (Roberts & Solow 2003; Solow 2005), (iii) Robson (Robson & Whitlock 1964), (iv) Solow1993.eq2 (Solow 1993a), (v) Solow2005.eq7 (Solow 2005), (vi) Strauss (Strauss & Sadler 1989). Very high numbers of sightings events caused the Burgman technique to fail, and so the number of sightings was converted to presence/absence data (i.e. an individual had been observed or not at that time point) for use with this technique.

Of the methods included in the sExtinct package, three (Burgman, Solow1993.eq2 and Solow2005.eq7) calculate the probability that a species has gone extinct at a given point in time. For these methods the package tests the probability of extinction iteratively at each time point after the last sighting event, up until a given date (the "test.year", see "sExtinct" help files). The date of extinction is then calculated as the date at which the probability of a species persisting falls below the alpha value. Typically for real world data the test.year will be set to the current year (i.e. what is the probability that a species is extinct). Preliminary testing (not presented) suggested that the maximum extinction date for the microcosm system would not fall above day 300, and for the real–world extirpations not above the year 2200, so we set the test.year conservatively at day 400 for the experimental data and 2300 for the wild population data. For those methods that simple produce a point estimate of the date of extinction from a sighting record (OLE, Robson, Strauss) only estimates that were less than or equal to day 400 were included in the analysis. After the simulations had been run, we are able to show that the maximum estimated date of extinction for the microcosm data was day 225, and year 2061 for the real-world data, both well under the point at which extinctions were tested up to.

Simulated samplings were run on the experimental and wild population data enough times to provide 500 extinction estimates for each combination of search regime and search regularity. For the experimental data the simulations were run 950 times for each individual experimental population. This number was chosen because the constant search regime had a fixed number of search efforts (95, see above), and this was then repeated 10 times to generate a high number of extinction estimates. This was then mirrored in the increasing and decreasing regimes. In total, across the replicate populations, search regimes, search regularities, and number of sighting events, this produced 631,452 simulations where at least four sighting events were produced (and thus an extinction estimate could be made).

Simulations were run on the wild population data in the same way, but because of the (generally) low population abundances, lack of replicate populations, and short observation periods, there were far fewer occasions where four sighting events were produced. Consequently simulations were run 3,800 times for each combination of search regime, number of sighting events and search regularity (four times as many as the experimental simulation), except for decreasing search effort and irregular sampling, which was run 7,600 to produce sufficient numbers of estimates of extinction. In all this produced a total of 734,121 sets of more than four sighting events (approximately similar to the 631,452 produced by the simulations run on the experimental data).

The outcomes of the different methods for inferring extinction were compared across the different search regimes and experimental treatments used. To assess the accuracy of each method, error was calculated as the difference between the inferred date of extinction and the observed date of extinction. Because the number and temporal distribution of sighting events were determined by the search regimes, all analysis were carried out on a subset of the data; 500 randomly selected extinction estimates from each search regime for each of the six methods for inferring extinction (3,000 estimates for each method). The wild population data covered a range of population decline rates, and a range of time spans that those declines were monitored over. Because of these different observation periods, error of inferred dates of extinction was normalised by dividing it by the minimum time between observations (for most species this was 1 year, but for some such as the Common Mist Frog, this was only four weeks).

All simulations were carried out using the R statistical software (R Development Core Team 2013). In the main we assess the accuracy of estimates in terms of relative error (the distance from the inferred date of extinction to the observed date of extinction, split into overestimation and underestimation of the extinction date). We look at the frequency and magnitude of overestimation and underestimation when assessing the overall robustness of each method.

Results – experimental data

Effects of environmental change

The rate and direction of temperature change altered the rate at which populations declined (Fig. 1b); warmer treatments produced faster rates of extinction, and cooler treatments slower. These environmentally driven rates of decline affected the accuracy of estimates, with mean error of estimates tending to be higher in cooler treatments, and lower in warmer treatments (Fig. 3). In general this effect was most noticeable in the change in the accuracy of underestimates of extinction, with all methods except Solow2005.eq7 showing a decrease in the mean error of underestimates at warmer temperatures.

The proportion of underestimates to overestimates of the extinction date was also affected by the treatments, with, in general, warmer treatments having a greater proportion of overestimates than cooler treatments, although the opposite was true for Solow1993.eq2 and Solow2005.eq7.

Effects of search regime

The effects of search regime and number of sighting events used (below) were calculated across data from the seven temperature treatments where extinction occurred. Search regime dramatically altered the accuracy of estimates (Fig. 4). For half of the methods (OLE, Solow1993.eq2, Strauss) error was minimised when the search regime was either constant or increasing, and the greatest error was generated when the search regime was decreasing (Fig. 4). For OLE and Solow1993.eq2 the vast majority of the error generated by decreasing search effort was underestimates of the date of extinction. Solow2005.eq7 produced no estimates of extinction when search efforts were increasing.

Burgman, Robson, and Strauss all showed similar patterns of error, with the greatest magnitude of overestimates occurring when the search regime was either constant or increasing (this error was typically much greater than OLE, Solow1993.eq2 or Solow2005.eq7), and the greatest magnitude of underestimates occurring when search effort was decreasing.

Effects of number of sightings used

The number of sighting events (time points at which sightings were recorded) used to infer extinction altered the accuracy of all of the methods tested (Fig. 5). In general the more sighting events used, the lower the mean error, this was especially true for underestimates of the date of extinction, which, across all methods, increased in accuracy as the number of sighting events used increased. In general, the greatest accuracy of estimates was achieved when the number of sighting events was greater than 10, and this was especially noticeable with OLE, Solow1993.eq2, and Solow2005.eq7. The Robson method was excluded from this analysis, as it uses only the last two recorded sighting events to estimate extinction.

Robustness of methods

The method used had a large impact on the accuracy of estimates (Fig. 6a). Mean absolute error (mean error normalized to positive values) was calculated for each method across all search regimes, and temperature treatments, to give an indication of each method’s applicability to real-world data (where information of search effort and rates of extinction are usually unknown). OLE produced the lowest mean error (7.9 days), with Solow2005.eq7 also having relatively low error (9.1 days). All other methods produced mean errors >10.4 days, with the greatest mean error associated with estimates made using Burgman (19.1 days). All methods, except Solow2005.eq7, inferred extinction to have occurred between day 0 and day 400, the last possible extinction date was day 70, in a high proportion of simulations (>0.999, Fig. 6a). OLE, Solow1993.eq2 and Solow2005.eq7 all produced less mean error than when a random method was selected for each inference of extinction, but more mean error than when method that produced the lowest error for each inference of extinction was selected (Fig. 6a).

When positive and negative errors are plotted separately, instead of being normalised to positive values, it becomes clear that most methods are prone to either overestimation or underestimation of the date of extinction (Fig. 6b). In some cases this bias is dramatic: Solow1993.eq2 and Solow2005.eq7 underestimate the date of extinction 99% and 92% of the time respectively, whilst Strauss, Robson and Burgman all appear to overestimate extinction (infer extinction to have occurred after it has already happened) more than 79% of the time (Fig. 6b). Only OLE shows little bias in the frequency of overestimation to underestimation. The magnitude of these errors is highly dependent on the method; however, in most cases the magnitude of error is consistently weighted to either underestimation or overestimation, with the exception of Robson, which is roughly evenly distributed (Fig. 6b). In many instances the difference in the magnitude of the mean error is large, for example OLE tends to have greater error when the estimate is an underestimate, rather than when it is an overestimate.

Results – wild population data

Effect of rate of population decline

The decline dynamics of the species altered the accuracy of inferred dates of extinction, but the relationship between the rate of decline and accuracy was dependent on the method used (Fig. 7). A general pattern of decreased accuracy with faster rate of population decline is apparent in all methods except Robson, which showed a decrease in mean error as populations declined at faster rates (Fig. 7). For OLE, Solow1993.eq2 and Solow2005.eq7 this pattern was driven in the main by an increase in the magnitude of the error associated with underestimates of extinction, whilst for Burgman the opposite is true (Fig. 7). Strauss showed an increase in the magnitude of the error associated with both overestimates and underestimates of extinction as populations declined more rapidly. Interestingly OLE, Solow1993.eq2, Solow2005.eq7 and Strauss all show very similar patterns of error across the different species.

Some species had consistently large error associated with their inferred dates of extinction across the majority of the methods tested (notably the Waterfall Frog, which tended to have an inferred extinction date significantly and consistently after the actual date of extinction, Fig. 7). No species had consistently low errors estimate error, although the Corncrake and Hawaiian Crow had low error in all estimates save those made by Robson (Fig. 7).

Discussion

We show that the rate at which a population has declined may influence the accuracy with which we can infer when that population has gone extinct. Previously it has been suggested that more rapid rates of decline may facilitate accurate inference of extinction (Rivadeneira et al. 2009; Clements et al. 2013), and this is indeed seen with some inference methods using data generated from microcosm communities. However, when using data from wild populations the opposite is often observed, with species that decline slowly typically having less error associated with inferred dates of extinction. In line with previous studies (Rivadeneira et al. 2009; Clements et al. 2013), we find that the search regime can strongly influence the accuracy of estimates, but that most important appears to be the inference method used, and that in general OLE (Solow 2005) is the most accurate and potentially most widely applicable of the methods tested.

The rate at which populations decline to extinction may vary based on generation time and reproductive output, as well as rates of biotic and abiotic environmental change (Fig. 1b, 2). We show that, whilst different rates of population decline can alter the accuracy of estimates, the nature of this effect is not necessarily consistent across different sources of data (Fig. 3, 7). Data from experimental populations produces results similar to those previously observed (Clements *et al.* 2013); more rapid rates of population decline, driven by environmental change, lead to more accurate estimates of extinction time. However, this was dependent on the identity of the method used, probably due to the different assumptions underlying each of the methods (Appendix S2), and consequently how each method predicts the probability of extinction changes through time (Appendix S5). For example, Solow2005.eq2 does not predict extinction where search efforts have increased over time, probably because the method assumes the pre-extinction sighting rate decreases, an assumption clearly broken when search effort increases over time. Burgman on the other hand consistently overestimates extinction when search efforts are increasing or constant, however this is likely to be in part due to the need to reduce high numbers of sighting to presence and absence data. When there are low numbers of sightings at each time point this method may perform better than is suggested by the results here.

It seems unlikely that such results will be particular only to extinctions driven by directional environmental change (as in the microcosms), as previous work using data from modeled populations has shown similar findings (Rivadeneira et al. 2009). Data from wild population extirpations, however, often show an opposing pattern, with increasing error of estimates when the rate of population decline was rapid (Fig. 3). The differing results generated using microcosm data and those from real world population extirpations may appear conflicting. This could be driven by the difference between the drivers of extinction (directional environmental change in the experimental set up and a variety of pressures in the wild population data), but is more likely to be driven by an interaction between the way sighting events are produced, and the (generally) lower abundances of the wild populations over short observation periods (Fig. 2, Appendix S3). When population abundances are low, and observation periods are short (e.g. the Corncrake, Fig. 2), there are only a limited number of possible times at which sighting events can be produced. This means that, unlike the microcosm data, there is limited time over which wild populations can produce temporally sporadic sightings. Sighting events are further reduced by decreasing search efforts, irregular sampling, and because some methods require at least four sighting events are required to infer extinction. Given that widely temporally spaced sighting events tend to produce estimates long after a population has been observed to go extinct, there are fewer opportunities for poor estimates of extinction to be produced. Consequently, when there are a small number of sighting events that are closely clustered the inferred date of extinction cannot fall far from the observed extinction event, a different scenario to when the only the most recent sighting events are used, as these can be widely spread in time (Fig. 5). This highlights a problem found in many records of contemporary wild population extirpations: they are both spatially and temporally limited. However, real historic sighting events may cover relatively long periods of time, with potentially a relatively high number of sighting events, a case in point being the most recent sightings of the dodo: 1598, 1601, 1602, 1607, 1611, 1628, 1628, 1631, 1638, 1662 (Roberts & Solow 2003). Consequently, we suggest that data from microcosm experiments may in fact be far more suitable for testing methods of inferring extinction, not only not only because the date of extinction can be accurately gauged (Clements et al. 2013), but because sighting records more akin to those found historically can be produced than are feasible using short abundance data sets from wild populations. Sighting records produced using microcosm data must then be compared to those typically found in real-world scenarios to see whether such sighting records are appropriate. When designing future microcosm-based experiments the conditions should be such that lower population abundances through time are produced to more accurately reflect wild populations declines, achieved by using lower temperatures, lower nutrient levels, or smaller habitats.

The rate and form with which a population declines, and historic search efforts, are both significant drivers of the temporal distribution of sighting events, and, consequently, both are important factors in determining the accuracy of inferred dates of extinction (Rivadeneira et al. 2009; Collen et al. 2010; Clements et al. 2013). To illustrate this conceptually, imagine a situation where a population declines slowly to extinction, but search efforts slowly increase, potentially due to increasing concern for that species; a constant frequency of sighting events could result, whilst masking the decline of a population up until an abrupt extinction event. Conversely, populations that crash from high abundances to extinction over a very short time period (e.g. *Euphydryas editha*, Thomas et al. 1996) may have high numbers of sighting events prior to extinction. However, sighting records are typically produced by sporadic chance observations of a species, often as a byproduct of some other endeavor, rather than systematic searches for an endangered species (Roberts & Solow 2003; though see Turvey et al. 2007). Gauging historic search efforts is therefore likely to be difficult. In real world terms, this means that appropriate choice of which method to apply, and the number of sighting events to use, are likely to be the two main ways in which error can be minimized. Identifying techniques that provide robust, accurate estimates over a variety of different potential drivers of error is thus of critical importance.

We find the method that produces the lowest mean error among our tests is OLE (Roberts & Solow 2003; Solow 2005), and that this pattern holds for both the experimental and wild population data (Fig. 6, Appendix S3). It also exhibits little bias towards either overestimating, or underestimating the date of extinction, although error in underestimates tends to be larger than that associated with overestimates. In addition, and unlike some other methods (notably Solow2005.eq2), OLE infers extinction to have occurred in a high proportion of the simulations (Fig. 6a, S3). This means that for many real-world situations, where historic search efforts and rates of population decline remain unknown, OLE should be regarded as the most reliable of the six methods tested here. Of particular importance to the read-world application of this method is our finding that using OLE with ten or more sighting events typically produces the most accurate estimates of extinction. This contradicts the widely held belief that OLE should be used with the 5 most recent sighting events only (Solow 2005), and necessitates a shift in how this method should be used in the future.

In situations where the search effort decreases through time, OLE (and in fact the majority of methods tested here) does poorly (Fig. 4). This is probably a function of infrequent sighting events that are not representative of actual population declines (e.g. sighting frequency declines rapidly, driven by search effort rather than population declines), and this pattern is often exacerbated when sampling is infrequent. The Robson and Strauss methods are the exception to this rule, as both inferred extinction in a high proportion of simulations, whilst performing better than the other four methods tested when search effort is decreasing and sampling is either regular or irregular (Fig. 4, 6a, Appendix S1). This greater accuracy is almost certainly cause by their tendency to overestimate the date of extinction in most other circumstances (Fig. 6b), making them less appropriate for use where search efforts are constant or increasing (Fig. 4). If there was some indication that the search effort through time that accompanied a series of historic sighting events had declined, then choosing either Robson or Strauss as an alternative to OLE could be appropriate. Where a more detailed knowledge of sampling intensity over time is known, other methods may be more appropriate than those tested here, for example search effort through time may be explicitly accounted for in the methods proposed by Marshall (1997) and McCarthy (1998). Such methods have been show to perform well where there are reasonable estimates of search intensity (Rivadeneira et al. 2009).

However, the availability of information on historic search efforts is often lacking, a function of the stochastic nature of sighting events, and potential solutions for effectively selecting extinction estimators in the absence of this information have previously been suggested (Vogel et al. 2009). For example the use of L-moment diagrams to assess how well the assumptions of each method are met by the underlying distribution of historic sighting events could be implemented (Vogel et al.

2009). Testing L-moment approach using experimental data with known extinction dates, and varying rates of species decline, could form an interesting future direction for the selection of such sighting based methods of extinction.

In conclusion, to accurately gauge the current rate of biodiversity loss we must be able to reliably classify a species as either extinct or extant, however many factors may influence our ability to infer extinction status correctly, not least the choice of inference method. In an ideal situation methods could be selected based on their strengths and weaknesses. Unfortunately, this is probably an unrealistic scenario given the often-poor knowledge of important factors such as search effort and rate of population decline. Consequently, methods should be applied that are robust to a variety of drivers of uncertainty. This work shows that in the majority of cases OLE (Roberts & Solow 2003; Solow 2005) provides the most accurate estimates of the extinction of experimental and wild populations. Importantly, and contrary to previous work (Solow 2005), we show that the accuracy of OLE improves as the number of sighting events used increases, and that ideally one should infer extinction using this technique with a minimum of 10 sighting records. Using such a robust technique will allow more accurate inference of the current extinction status of species than would be possible if one were to pick one of the six methods tested here without any prior knowledge. However, in certain circumstances (especially when historic search efforts have been decreasing and searching has been irregular) inferred dates of extinction should be treated with care. If there was an indication that this had occurred, using either Robson (Robson & Whitlock 1964) or Strauss (Strauss & Sadler 1989) instead could be appropriate. Where greater information on search efforts is available, techniques that explicitly account for search intensity should be considered (e.g. Marshall 1997; McCarthy 1998).

Acknowledgements

This research was funded by a Natural Environment Research Council CASE studentship grant in partnership with the Zoological Society of London (NE/H018700/1). B.Fagan facilitated access to datasets that he had previously compiled from the literature. We also thank the input of the editors and referees whose comments have significantly improved the quality of this manuscript.

Supporting Information

Appendices S1, S2, S3, S4, and S5 are available online. The authors are solely responsible for the content and functionality of these materials. Queries (other than absence of the material) should be directed to the corresponding author.

References

Barnosky, A. D. et al. 2011. Has the Earth's sixth mass extinction already arrived? Nature **471**:51–57. Nature Publishing Group.

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

Burrows, R., H. Hofer, and M. L. East. 1995. Population dynamics, intervention and survival in African wild dogs (Lycaon pictus). Proceedings of the Royal Society B: Biological Sciences **262**:235–245. The Royal Society.

Clements, C. F., N. Worsfold, P. Warren, B. Collen, T. Blackburn, N. Clark, and O. L. Petchey. 2013. Experimentally testing an extinction estimator: Solow's Optimal Linear Estimation model. Journal of Animal Ecology **82**:345–354.

Collen, B., J. Loh, S. Whitmee, L. McRae, R. Amin, and J. E. M. Baillie. 2009. Monitoring change in vertebrate abundance: the living planet index. Conservation Biology **23**:317–27. Blackwell Publishing Inc.

Collen, B., A. Purvis, and G. M. Mace. 2010. When is a species really extinct? Testing extinction inference from a sighting record to inform conservation assessment. Diversity and Distributions **16**:755–764.

Crowley, T. J. 2000. Causes of climate change over the past 1000 years. Science **289**:270–277. AAAS.

Di Fonzo, M., B. Collen, and G. M. Mace. 2013. A new method for identifying rapid decline dynamics in wild vertebrate populations. Ecology and Evolution.

Fagan, W. F., and E. E. Holmes. 2006. Quantifying the extinction vortex. Ecology Letters **9**:51–60.

Fisher, D. O., and S. P. Blomberg. 2011. Correlates of rediscovery and the detectability of extinction in mammals. Proceedings of the Royal Society of London B **278**:1090–1097.

Gotelli, N. J., A. Chao, R. K. Colwell, W.-H. Hwang, and G. R. Graves. 2011. Specimen-based modeling, stopping rules, and the extinction of the Ivory-Billed Woodpecker. Conservation Biology **26**:47–56.

Laurance, W. F., K. R. McDonald, and R. Speare. 1996. Epidemic disease and the catastrophic decline of Australian rain forest frogs. Conservation Biology **10**:406–413. Blackwell Scientific Publications.

Lawler, S. P., and P. J. Morin. 1993. Food-web architecture and population-dynamics in laboratory microcosms of protists. The American Naturalist **141**:675–686.

Marshall, C. R. 1997. Confidence intervals on stratigraphic ranges with nonrandom distributions of fossil horizons. Paleobiology **23**:165–173. Paleontological Society. Retrieved from http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-0030859884&partnerID=40&rel=R7.0.0.

McCarthy, M. A. 1998. Identifying declining and threatened species with museum data. Biological Conservation **83**:9–17. Elsevier.

McPherson, J. M., and R. A. Myers. 2009. How to infer population trends in sparse data: examples with opportunistic sighting records for great white sharks. Diversity and Distributions **15**:880–890. WILEY-BLACKWELL PUBLISHING.

Parr, R. 1992. The decline to extinction of a population of Golden Plover in north-east Scotland. Ornis Scandinavica **23**:152–158.

R Development Core Team. 2013. R: A language and environment for statistical computing. Retrieved from http://www.r-project.org/.

Rivadeneira, M. M., G. Hunt, and K. Roy. 2009. The use of sighting records to infer species extinctions: an evaluation of different methods. Ecology **90**:1291–1300.

Roberts, D. L., and A. R. Solow. 2003. Flightless birds: When did the dodo become extinct? Nature **426**:245.

Robson, D. S., and J. H. Whitlock. 1964. Estimation of a truncation point. Biometrika **31**:33–39.

Solow. 1993a. Inferring extinction from sighting data. Ecology **74**:962–964. JSTOR.

Solow. 2005. Inferring extinction from a sighting record. Mathematical Biosciences **195**:47–55.

Solow, A. R. 1993b. Inferring extinction in a declining population. Journal of Mathematical Biology **32**:79–82. Springer.

Solow, A. R., and D. L. Roberts. 2003. A nonparametric test for extinction based on a sighting record. Ecology **84**:1329–1332.

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.

Thomas, C. D., M. C. Singer, and D. A. Boughton. 1996. Catastrophic extinction of population sources in a butterfly metapopulation. The American Naturalist **148**:957–975. The University of Chicago Press.

Thompson, A., J. Rhodes, and I. Pettman. 1988. Culture Collection of Algae and Protozoa (CCAP), Freshwater Biological Association, Ambleside, Cumbria, UK. Titus Wilson & Son Ltd.

Turvey, S. T. et al. 2007. First human-caused extinction of a cetacean species? Biology Letters **3**:537–540. The Royal Society.

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. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19198956.

Weimerskirch, H., and P. Jouventin. 1987. Population Dynamics of the Wandering Albatross, Diomedea exulans, of the Crozet Islands: Causes and Consequences of the Population Decline. Oikos **49**:315–322.