Year: 2014

Effects of recent environmental change on accuracy of inferences of extinction status

Clements, Christopher F ; Collen, Ben ; Blackburn, Tim M ; Petchey, Owen L

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, results of recent experiments with microcosm communities suggest that the rate at which a population declines to extinction, potentially driven by varying environmental conditions, may alter one's ability accurately to infer extinction status. We tested how the rate of population decline, driven by historic environmental change, alters the accuracy of 6 commonly applied sighting-based methods used to infer extinction. We used data from small-scale experimental communities and recorded wild population extirpations. We assessed how accuracy of the different methods was affected by rate of population decline, search effort, and number of sighting events recorded. Rate of population decline and historic population size of the species affected the accuracy of inferred extinction dates; however, 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. 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 so that they mirror closely the abundance dynamics of populations affected by real-world extirpation events.

Title: Effects of Recent environmental change on accuracy of inferences of extinction status

Running head: *Accurately inferring extinction status*

Word count: 6,170; your response-to-edits manuscript should not exceed 6200 words

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

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

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, results of recent experiments with microcosm communities suggest that the rate at which a population declines to extinction, potentially driven by varying environmental conditions, may alter one's 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 used to infer extinction. We used data from small-scale experimental communities and recorded wild population extirpations. We assessed how accuracy of the different methods was affected by rate of population decline, search effort, and number of sighting events recorded. Rate of population decline and historic population size of the species affected the accuracy of inferred extinction dates; however, faster declines produced more accurate inferred dates of extinction, but only when population sizes were higher. Optimal linear estimation 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. Optimal linear estimation 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 so that they mirror closely the abundance dynamics of populations affected by 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. 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 (Burgman et al. 1995; McCarthy 1998; Roberts & Solow 2003; Solow 2005). 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 1993, 2005; Roberts & Solow 2003; Solow & Roberts 2003; McPherson & Myers 2009). Recent work suggests 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 used experimental microcosm communities to provide detailed abundance time series data for species for which the date of extinction was 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; Clements et al. 2013). Where species persist at low density for a lengthy period (and thus are rarely observed), estimates are worse than when the species falls rapidly to extinction. Thus, historic pressures on a species (be they abiotic, such as temperature change or habitat loss, or biotic, such as an invasive species or disease) that increase the rate at which

4

a species declines may alter one's ability to judge accurately whether the species has in fact been lost and over what time frame extinction may have occurred.

Data collected on wild populations show that both the identity of the species (and thus life history) and 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) found distinct differences in the rates of rapid disease driven population declines of four species of rain forest dwelling frogs, and Weimerskirch and Jouventin (1987) recorded differences in the rates of population decline of the Wandering Albatross (*Diomedea exulans*) 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 one's ability to infer whether a species is extinct is of concern, and quantifying this effect is an issue that may affect our current understanding of the scale of biodiversity loss.

We used small-scale experimental communities to test whether there was a negative correlation between the rate at which a population declined 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 were subjected to different rates of environmental change, which altered their rate of 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 a time series of the abundance of these experimental and wild populations, we used three search regimes to generate a time series of sighting events (Rivadeneira et al. 2009; Clements et al. 2013). We then examined the effect of the rate of population decline , search regime, and number of

5

sighting events the calculations used on the accuracy of inferred dates of extinction and compared the robustness of the six techniques. On the basis of our results, we make recommendations about the use of such techniques in real-world scenarios.

Methods

Experimental and Wild population data

Experimental data were gathered on the abundances of replicate populations of the bactiverous ciliate *Loxocephalus* sp., which were subjected to various rates of temperature change through time (Supporting Information). These different rates of directional temperature change produced varying rates of population decline and thus times of extinction (Fig. 1a). Detailed explanations of the experimental set up, temperature treatments, and sampling method are described in Clements et al. 2014) and Supporting Information.

Data on population dynamics and extirpation events were collected from the literature (Google Scholar search for terms such as *extirpation*, *population extinction*, and *extinction dynamics*), the Living Planet Database (Collen et al. 2009), and Fagan and Holmes (2006). From these data sets, we selected eight time series (Parr 1992; Burrows et al. 1995; Laurance et al. 1996; Fagan & Holmes 2006) for African wild dog (*Lycaon pictus*), Hawaiian Crow (*Corvus hawaiiensis*), Corncrake (*Crex crex*), Whooping Crane (*Grus Americana*), European Golden Plover (*Pluvialis apricaria*), and waterfall frog (*Litoria nannotis*), common mist frog (*Litoria rheocola*), and sharp snouted day frog (*Taudactylus acutirostris*). Each species had 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. 1b) that are hypothesised to have been caused by a variety of factors including extreme weather events, disease, habitat loss and degradation, and invasive species. These rates of decline ranged

6

from very rapid (e.g. common mist frog), where approximately 40% of the initial population was lost per year, to relatively slow (e.g. 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) (methods). To these records of abundance through time three simulated search regimes were applied: constant, increasing, and decreasing effort. The constant search regime was simulated with search efforts (the fraction of the habitat searched) 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 that increased by a random fraction at each time step until the search effort reached 0.95, after which it remained constant. The decreasing search effort mirrored the increasing search effort: the fraction of habitat decreased through time. These simulated search efforts at each point were then used to generate series of sighting events. Multiplying the search effort (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 because sampling of wildlife populations is often sporadic. Thus, two search regularities were simulated: regular sampling (as above) and irregular sampling. Irregular sampling was implemented identically to regular sampling, except that with irregular sampling there was a 50% probability of every point where observations occurred being used. All analyses present

results that include data from both regular and irregular sampling. Results of the effect of regularity of sampling on the accuracy of extinction estimates are in Supporting Information.

Hereafter, we refer to the times at which sightings were recorded as sighting events; that is, days in the experimental system and months or years in the wild population data. At each of these points there are a number of observations (referred to as sightings) generated, which depend on the abundance of the population and the amount of the habitat searched. If there were fewer than four sighting events, then estimates were not made.

The sighting records were used to test the six sighting-based methods for inferring historical extinction currently included in the R package sExtinct (Supporting Information). These methods do not explicitly take into account the search effort that generated a sighting event, although methods that do this exist (e.g., Marshall [1997]; McCarthy [1998]). Thus, given that the methods we used 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 are referred to by simplifications of the function names in the R package, as follows: Burgman (Burgman et al. 1995), optimal linear estimation (OLE) (Roberts & Solow 2003; Solow 2005), Robson (Robson & Whitlock 1964), Solow1993.eq2 (Solow 1993), Solow2005.eq7 (Solow 2005), and Strauss (Strauss & Sadler 1989). Very high numbers of sighting events caused the Burgman technique to fail, so we converted the number of sightings to presence-absence data (i.e., an individual was observed or was not observed at that point) for use with this technique. It should be noted that the techniques tested here do not estimate the time at which a species has gone extinct. Rather, they infer the time at which a species can be declared extinct (i.e. probabilistically the species will have gone extinct prior to this inferred date).

Of the methods included in the sExtinct package, three (Burgman, Solow1993.eq2, and Solow2005.eq7) calculate the probability that a species had gone extinct by a given

point. For these methods, the package tests the probability of extinction iteratively at each point after the last sighting event up to a given date (the test.year, see sExtinct help files). The date at which extinction can be declared is then calculated as the date at which the probability given by the model 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 be before day 300, and for the real–world extirpations it would not be before year 2200. So, we set the test.year conservatively at day 400 for the experimental data and at year 2300 for the wild population data. For those methods that simply produce a point estimate of the date of extinction from a sighting record (OLE, Robson, Strauss), only estimates that ≤day 400 were included in the analysis. After the simulations had been run, we found that the maximum estimated date of extinction for the microcosm data was day 225 and was year 2061 for the real-world data, both well under the time up to which extinctions were tested.

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), and the simulation was repeated 10 times to generate a high number of extinction estimates. This step 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 procedure produced 631,452 simulations where at least four sighting events were produced (and thus an extinction estimate could be made). Given the generally low population sizes and short time series of the wild population data, all simulations were repeated four times as often as the experimental simulations to

produce a sufficient number of occasions where four sighting events occurred, producing a total of 746,148 sets of more than four sighting events.

To assess the accuracy of each method, we calculated error 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. The wild population data covered a range of population decline rates and a range of time spans over which those declines were monitored. Because of these different observation periods, error of inferred dates of extinction was normalized by dividing it by the minimum time between observations (for most species this was 1 year, but it could be as little as 4 weeks [e.g. for common mist frogs]).The effects of search regime and number of sighting events used (below) were calculated across data from the seven temperature treatments where extinction occurred.

To assess the robustness of the methods, 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).

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

Results

Experimental data

10

The rate and direction of temperature change altered the rate at which populations declined (Fig. 1a): warmer treatments produced faster rates of extinction, and cooler treatments produced slower rates of extinction. These environmentally driven rates of decline affected the accuracy of estimates; mean error of estimates tended to be higher in cooler treatments and lower in warmer treatments (Fig. 2). In general this effect was most noticeable in the change in the accuracy of underestimates of extinction; all methods showed 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. In general warmer treatments had a greater proportion of overestimates than cooler treatments, although this was not true for Solow1993.eq2 and Solow2005.eq7.

Search regime dramatically altered the accuracy of estimates (Fig. 3). For half of the methods (OLE, Solow1993.eq2, Strauss), error was minimized when the search regime was either constant or increasing, and the greatest error was generated when the search regime was decreasing (Fig. 3). 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 for Burgman, Robson, and Strauss than for OLE, Solow1993.eq2, or Solow2005.eq7) and the greatest magnitude of underestimates occurred when search effort was decreasing.

The number of sighting events (points at which sightings were recorded) used to infer extinction altered the accuracy of all the methods tested (Fig. 4). 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 >10, and this was especially noticeable with OLE, Solow1993.eq2, and Solow2005.eq7. The Robson method was excluded from this analysis because it uses only the last two recorded sighting events to estimate extinction.

The method used had a large impact on the accuracy of estimates (Fig. 5a). The OLE method produced the lowest mean error (7.9 days); Solow2005.eq7 also had relatively low error (9.1 days). All other methods produced mean errors >10.4 days, and the greatest mean error was associated with estimates made with Burgman (19.1 days). All methods inferred extinction to have occurred from day 0 to day 400 (last observed extinction date was day 70) in a high proportion of simulations, except Solow2005.eq7, where extinction was inferred to have occurred in only one-quarter of the simulations (Fig. 5a). Solow1993.eq2 , Solow2005.eq7, and OLE all produced less mean error than when a random method was selected for each inference of extinction, but they produced more mean error than when the method that produced the lowest error for each inference of extinction was selected (Fig. 5a).

When positive and negative errors are plotted separately, instead of being normalized to positive values, it became clear that most methods were prone to either overestimation or underestimation of the date of extinction (Fig. 5b). In some cases, this bias was dramatic. Solow1993.eq2 and Solow2005.eq7 underestimated the date of extinction 99% and 95% of the time respectively, whilst Strauss, Robson, and Burgman all appeared to overestimate extinction over 79% of the time (Fig. 5b). Only OLE showed little bias in the frequency of overestimation to underestimation. The magnitude of these errors was highly dependent on

the method; however, in most cases the magnitude of error was consistently weighted to either underestimation or overestimation, with the exception of Robson, which was roughly evenly distributed (Fig. 5b). In many instances, the difference in the magnitude of the mean error was large. For example, OLE tended to have greater error when the estimate was an underestimate, rather than when it is an overestimate.

Wild population data

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. 6). A general pattern of decreased accuracy with faster rate of population decline was apparent in all methods except Robson, which showed a decrease in mean error as populations declined at faster rates (Fig. 6). 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 was true (Fig. 6). 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 showed 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. 6). No species had consistently low error estimates, although the Corncrake and Hawaiian Crow had low error in all estimates save those made by Robson (Fig. 6).

Discussion

We found that the rate at which a population declined affected the accuracy with which one could infer when that population became extinct. 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 was indeed seen with some inference methods in which data generated from microcosm communities were used. However, when using data from wild populations, the opposite is often observed: species that decline slowly typically have less error associated with inferred dates of extinction. In line with previous studies (Rivadeneira et al. 2009; Clements et al. 2013), we found 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 varied based on generation time, reproductive output, and rates of biotic and abiotic environmental change (Fig. 1a, b). We found that whilst different rates of population decline can alter the accuracy of estimates, the nature of this effect was not necessarily consistent across different sources of data (Fig. 2, 6). Data from experimental populations produced 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 outcome depended on the method used, probably due to the different assumptions underlying each of the methods (Supporting Information). Consequently, how each method inferred the probability of extinction changed through time (Supporting Information). For example, Solow2005.eq2 did not predict extinction where search effort 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 consistently overestimated extinction when search efforts were increasing or constant; however, this was likely in part due to the need to

14

reduce high numbers of sightings to presence and absence data. When there were low numbers of multiple sightings at each point this method may perform better than is suggested by our results.

It seems unlikely that such results will be particular only to extinctions driven by directional environmental change (as in the microcosms) because previous work with data from modeled populations has shown similar findings (Rivadeneira et al. 2009). Data from wild population extirpations, however, often showed an opposing pattern, with increasing error of estimates when the rate of population decline was rapid (Fig. 2). The differing results generated using microcosm data and 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 it is more likely to be driven by an interaction between the way sighting events are produced and the (generally) lower abundances of wild populations over short observation periods (Fig. 1b, Supporting Information). When population abundances were low and observation periods were short (e.g. the Corncrake, Fig. 1b), there were only a limited number of possible times at which sighting events could be produced. This means that, unlike the microcosm data, there was limited time over which wild populations could produce temporally sporadic sightings. Sighting events were further reduced by decreasing search efforts and irregular sampling and because some methods require at least four sighting events 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 from when only the most recent sighting events were used because these can be widely spread in time (Fig. 4).

15

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 and potentially have a relatively high number of sighting events, a case in point being the most recent sightings of the Dodo (*Raphus cucullatus T.*): 1598, 1601, 1602, 1607, 1611, 1628, 1628, 1631, 1638, 1662 (Roberts & Solow 2003). Consequently, we suggest that data from microcosm experiments may be far more suitable for testing methods of inferring extinction not only because the date of extinction can be accurately gauged (Clements et al. 2013), but also 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 more accurately to reflect wild population declines, which can be achieved by using lower temperatures, lower nutrient levels, or smaller habitats.

Rates at which and forms of population declines and historic search efforts are significant drivers of the temporal distribution of sighting events. 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 that could mask the decline of a population up until an abrupt extinction event. Conversely, populations that crash from high abundances to extinction over a short 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). Gauging historic search efforts is therefore likely to be difficult. 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 found that OLE produced the lowest mean error among our tests (Roberts & Solow 2003) (Fig. 5, Supporting Information) and that it exhibited relatively little bias toward either overestimating or underestimating the date of extinction. Unlike some other methods (notably Solow2005.eq2), OLE inferred that extinction occurred in a high proportion of the simulations (Fig. 5a, Supporting Information). 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 we tested. Of particular importance to the real-world application of this method is our finding that used of OLE with ten or more sighting events typically produced 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). Consequently, we recommend a change in how this method is used in the future.

In situations where the search effort decreases through time, OLE, and in fact the majority of methods tested here, performed poorly (Fig. 3). This is probably a function of infrequent sighting events that were not representative of actual population declines (e.g. sighting frequency declines rapidly driven by search effort rather than population declines). Irregular sampling often exacerbates this pattern, although the size of this effect is rather small and in general the methods tested are robust to the regularity of sampling, an encouraging finding when applying such methods to real-world data (Supporting

17

Information). The Robson and Strauss methods performed better than the other four methods tested when search effort was decreasing (Fig. 3), almost certainly a function of their tendency to overestimate the date of extinction in most other circumstances (Fig. 5b), a fact that makes them less appropriate for use where search efforts are constant or increasing (Fig. 3). If there were an indication that the search effort through time that had produced a series of historic sighting events had declined but the exact search effort had not been recorded, then choosing either Robson or Strauss as an alternative to OLE would be appropriate. However, if no such information is available, then OLE should be used because it is relatively robust to search effort and overall produces accurate estimates of extinction.

In situations where a more detailed knowledge of sampling intensity over time is known, other methods may be applicable. For example, search effort through time may be explicitly accounted for in the methods proposed by Marshall (1997) and McCarthy (1998). Such methods 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 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 the 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.

To gauge accurately the current rate of biodiversity loss, one must be able to reliably classify a species as either extinct or extant. However, many factors may influence one's 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. Unfortunately, this is probably an unrealistic scenario given the often poor knowledge of important factors that relate to these strengths, such as search effort and rate of population decline. Consequently, methods should be applied that are robust to a variety of drivers of uncertainty. Our results show that in the majority of cases OLE (Roberts & Solow 2003; Solow 2005) provided the most accurate estimates of the extinction of experimental and wild populations. Importantly, and contrary to previous work (Solow 2005), we found that the accuracy of OLE improved as the number of sighting events used increased and that ideally one should infer extinction with this technique with a minimum of ten sighting records. Use of 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 decreased over time) 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) 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).

Supporting Information

Experimental set up and sampling (Appendix S1), Effect of search regularity on the accuracy of extinction estimates (Appendix S2), Brief summaries of each of the methods (Appendix S3), Probability of extinction on a method-by-method basis (Appendix S4), Interaction between small population size and rate of decline on the accuracy of estimates (Appendix S5), and Overall accuracy of methods using wild population data (Appendix S6) 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.

Literature Cited

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

Burgman, M.A., Grimson, R.C. & Ferson, S. (1995). Inferring threat from scientific collections. *Conservation Biology*, **9**, 923–928.

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

Clements, C.F., Collen, B., Blackburn, T.M. & Petchey, O.L. (2014) Effects of directional environmental change on extinction dynamics in experimental microbial communities are predicted by a simple model. *Oikos*, 123, 141-150.

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

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

Collen, B., Purvis, A. & Mace, G.M. (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.

Fagan, W.F. & Holmes, E.E. (2006). Quantifying the extinction vortex. *Ecology Letters*, **9**, 51–60.

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

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

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

Marshall, C.R. (1997). Confidence intervals on stratigraphic ranges with nonrandom distributions of fossil horizons. *Paleobiology*, **23**, 165–173.

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

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

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

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

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

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

Solow. (1993). Inferring extinction from sighting data. *Ecology*, **74**, 962–964.

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

Strauss, D. & Sadler, P.M. (1989). Classical confidence intervals and Bayesian probability estimates for ends of local taxon ranges. *Mathematical Geology*, **21**, 411–427.

Thomas, C.D., Singer, M.C. & Boughton, D.A. (1996). Catastrophic extinction of population sources in a butterfly metapopulation. *The American Naturalist*, **148**, 957–975.

Vogel, R.M., Hosking, J.R.M., Elphick, C.S., Roberts, D.L. & Reed, J.M. (2009). Goodness of fit of probability distributions for sightings as species approach extinction. *Bulletin of Mathematical Biology*, **71**, 701–719.

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

Figure 1. (a) Means over time of abundance of
replicate populations (error bars not included to improve clarity) subjected to different
termperature regimes, b) the dynamics of 8 wild population extirpations collected from the
literature (shaded areas, 1 SE; slope, number in upper right of each graph)

Figure 2. Mean error of inferred dates of made with each method (Burgman, OLE,Robson,
Solow 1993, Solow 2005, Strauss) across the different temperature
treatments ordered from fastest decreasing on the left to fastest increasing on
the right (bars, 1 SE; numbers
above and below bars, proportion of estimates that were overestimates or
underestimates, respectively, of the date of extinction). Error is the difference between the
inferred date of extinction and the observed
date of extinction

Figure 3. Mean error of inferred extinction for each method (Burgman, OLE,Robson, Solow
1993, Solow 2005, Strauss) under the different simulated
search regimes (bars,
1 SE; numbers above and below bars are the proportion of estimates that were
overestimates and underestimates, respectively, of the date of extinction; con, constant; dec,
decreasing; inc, increasing). Error is the difference between the inferred date of extinction
and the
observed date of extinction

Figure 4. Mean error of estimates of extinction relative to the number of sighting events (i.e.,

dates at

which sightings occurred) made by each method across all search regimes and experimental

temperature

treatments (error bars, 1 SE; trends are shown by a LOESS smoothing; shaded areas,

1 SE of the LOESS). Error is the difference between the inferred date of extinction and the

observed

date of extinction. The effect of the number of sightings on the Robson

method is not presented because the technique estimates extinction from the distribution of

the

last 2 sighting events only.

Figure 5. (a) Mean absolute error and (b) mean relative error of between inferred date of

extinction and observed date of extinction made by each

method across all search regimes, temperature treatments, and number of sighting events

(bars

, 1 SE). Error is the difference between the inferred date of extinction and the

observed date of extinction. The mean error when a random method

is chosen for each inference of extinction and when the method with the lowest error is

always

chosen are also shown. In (a) numbers in bars are the fraction of the simulations where

there were 4 or more sighting events produced for which each method produced an extinction

estimate that was from 0 to 140 days (note that the last potential extinction event in the

experimental treatment was day 70, and that this is not shown for the lowest error because

selecting the method with the lowest error inherently means an estimate has been

made). In (b) numbers in bars are the proportion of those estimates that fall after the observed

date of

extinction (top number) and before the observed date of extinction (bottom number).

Figure 6. Effect of species identity on accuracy of inferred dates of extinction across all

search regimes and numbers of sighting events (bars,

1 SE; numbers above and below bars, proportion of estimates that

were overestimates or underestimates, respectively, of the date of extinction)

. Species are ordered by the rate of population decline,

from the slowest (on the left) to the fastest (on the right), with the expectation being that

populations that decline slowly are likely to have greater mean error than populations that

decline faster.