Experimentally testing the accuracy of an extinction estimator: Solow's optimal linear estimation model

Christopher F. Clements[1]*, Nicholas T. Worsfold[2], Philip H. Warren[1], Ben Collen[3], Nick Clark[1], Tim M. Blackburn[3,4] and Owen L. Petchey[5]

[1]*Department of Animal and Plant Sciences, University of Sheffield, Sheffield, S10 2TN, UK;* [2]*Department of Politics, University of York, York, YO10 5DD, UK;* [3]*Institute of Zoology, ZSL, Regent's Park, London, NW1 4RY, UK;* [4]*Distinguished Scientist Fellowship Program, King Saud University, P.O. Box 2455, Riyadh, 1145, Saudi Arabia; and* [5]*Institute of Evolutionary Biology and Environmental Studies, The University of Zurich, Zurich, CH-8057, Switzerland*

Summary

1. Mathematical methods for inferring time to extinction have been widely applied but poorly tested. Optimal linear estimation (also called the 'Weibull' or 'Weibull extreme value' model) infers time to extinction from a temporal distribution of species sightings. Previous studies have suggested optimal linear estimation provides accurate estimates of extinction time for some species; however, an in-depth test of the technique is lacking.

2. The use of data from wild populations to gauge the error associated with estimations is often limited by very approximate estimates of the actual extinction date and poor sighting records. Microcosms provide a system in which the accuracy of estimations can be tested against known extinction dates, whilst incorporating a variety of extinction rates created by changing environmental conditions, species identity and species richness.

3. We present the first use of experimental microcosm data to exhaustively test the accuracy of one sighting-based method of inferring time of extinction under a range of search efforts, search regimes, sighting frequencies and extinction rates.

4. Our results show that the accuracy of optimal linear estimation can be affected by both observer-controlled parameters, such as change in search effort, and inherent features of the system, such as species identity. Whilst optimal linear estimation provides generally accurate and precise estimates, the technique is susceptible to both overestimation and underestimation of extinction date.

5. Microcosm experiments provide a framework within which the accuracy of extinction predictors can be clearly gauged. Variables such as search effort, search regularity and species identity can significantly affect the accuracy of estimates and should be taken into account when testing extinction predictors in the future.

Key-words: extinction estimation, optimal linear estimation, protist microcosm, Solow model, Weibull

Introduction

Species are being lost at an unprecedented rate (Barnosky *et al*. 2011). This loss of biodiversity could have huge impacts both on ecosystem function and on the survival of co-dependent extant species (Memmott, Waser & Price 2004). Biodiversity loss is one measure used to infer human impacts on ecosystems; however, the exact rate of species loss is unclear, and the fate of many species remains unknown (e.g. Rös & Pineda 2009). This has been highlighted by the re-appearance of 36% of mammal species (67 species) classified as extinct since A.D.1500 (Fisher & Blomberg 2011), the so-called Lazarus effect (Keith & Burgman 2004).

Categorizing a species as extinct with confidence is surprisingly difficult, as searching an entire species range is often unfeasible, and detection probabilities may be low, for example, if the species is cryptic, small, has a large range or low population density (Solow 2005). Inferring extinction from data produced by less than exhaustive searching would reduce the effort spent attempting to

classify a species as extinct. Recent examples such as the search for *Campephilus principalis* (Fitzpatrick *et al.* 2005; Fitzpatrick 2006), the ivory billed woodpecker, highlight the advantage of quantitatively estimating the extinction status of a species, prior to a large investment of time and money in the search for any remaining individuals (Jackson 2006; Rout, Heinze & McCarthy 2010). Consequently, a considerable number of extinction predictors have been proposed over the last 20 years (e.g. Solow 1993a,b, 2005; Mccarthy 1998; Solow & Roberts 2003; McInerny *et al.* 2006; Gotelli *et al.* 2011). A large proportion of these aim to infer extinction from sighting only (as opposed to abundance) data in an attempt to deal with the many species for which only a few sightings have been recorded (Fisher & Blomberg 2012).

Clearly inferring the date of extinction of a species could be of considerable practical value, but if such techniques are to be widely applied to the critical question of extinction status, thorough testing to identify their limitations is essential (Collen, Purvis & Mace 2010). Gauging how well these predictors of time of extinction will cope under real-world conditions has historically been tackled by using either data from modelled populations or sighting data from species thought to be currently either extinct or on the verge of extinction (Rivadeneira, Hunt & Roy 2009; Collen, Purvis & Mace 2010; Elphick, Roberts & Reed 2010; Fisher & Blomberg 2011). Collen, Purvis & Mace (2010) applied 'optimal linear estimation' to assess the probability of extinction in 10 mammal and 10 bird species from Oceania and Asia, respectively. For those species with more than five reported sightings, the technique appeared to provide accurate predictions. However, species with few or widely spaced sightings were assigned extinction times with upper confidence intervals often millennia into the future, suggesting that the null hypothesis (that the species is extant) cannot be rejected, even when the last recorded sighting of the species was over 100 years ago. Rivadeneira, Hunt & Roy (2009) assessed the accuracy of several sighting-based estimators of time of extinction using data from model populations. Accuracy was defined as the date of extinction of the model population falling within the upper 95% confidence interval produced by each method. Predictions tended to be more accurate when populations fell rapidly (as opposed to gradually) to extinction. Unfortunately, given the often-wide confidence intervals, the use of confidence intervals as a metric of accuracy appears unhelpful (Collen, Purvis & Mace 2010). Fisher & Blomberg (2011) attempted to assess the accuracy of three sighting-only extinction estimators (stationary Poisson, nonparametric and the Weibull/optimal linear estimation) by comparing predicted probabilities of extinction with current IUCN classification, using data on mammal species currently thought to be extinct or possibly extinct. However, the inherent problem with using data from wild populations is that there are very few examples where the time of extinction has been accurately observed (Collen, Purvis &

Mace 2010). Data such as those published in Fagan & Holmes (2006), where species have been monitored for population abundances preceding a local extinction event, offers the opportunity to test extinction indices on extinctions in natural communities. Unfortunately, such data sets are few and far between, and often only deal with very small populations just prior to extinction. Given the unknown geographic range of such populations, the effects of varying the amount of a habitat searched (search effort), potentially one of the most important factors affecting our ability to accurately of infer the date of extinction, cannot readily be tested.

Small-scale experimental communities offer an opportunity to test specific methods for inferring extinction on data generated by living organisms, under a range of biotic and abiotic treatments. Given the replicability, ease of manipulation of a variety of environmental variables, short generation time and large number of readily culturable species, microcosms can be used to bridge the gap between modelled and field data. Where an experimental microcosm approach provides particular advantages in testing extinction estimators is the ability to search an entire habitat of known size and accurately assess whether a species is extinct or not. With regular sampling, population declines and extinction dates can be quantified to a level of accuracy that is simply not feasible in field systems.

Here, we concentrate on exhaustively testing a single method of inferring extinction date based on sighting only data – optimal linear estimation (hereafter OLE) (Roberts & Solow 2003; Solow 2005), and outline a new framework within which extinction predictors can be tested in the future. Our aim is not to compare multiple indices, which has been addressed previously (e.g. Rivadeneira, Hunt & Roy 2009), but to use the combination of experimental microcosm systems, with a specific method for predicting extinction, to investigate whether varying search efforts and population dynamics are likely to impact our ability to infer extinction. OLE has received attention for several reasons: (i) OLE does not assume that detection probability or abundance prior to extinction is stable (assumptions that are made by other models, specifically the stationary Poisson model) (Fisher & Blomberg 2012), (ii) it has been found to perform well even when a species is declining in abundance gradually through time, and when the probability of observing the species is low (Rivadeneira, Hunt & Roy 2009) and (iii) it has been proposed as a method to inform an IUCN category of 'Critically Endangered – Possibly Extinct' (Collen, Purvis & Mace 2010). Given the importance of IUCN categorization, and the potential perils of wrongly classifying a species as extinct (or extant), thorough testing of this method is critical.

Using a time series of historical sighting events, OLE can be used to infer the date of extinction of a species which is currently thought (or assumed) to be extinct (Roberts & Solow 2003; Solow 2005). Observations of an individual (or multiple individuals at the same point in

time) are recorded as a single sighting event: the species has been observed as being extant at a given point in time. OLE assumes that the joint distribution of the most recent sighting events has the form of a Weibull extreme value distribution and estimates the shape parameter of this distribution from the temporal spread of the sighting events. In a simple example, a vector of sighting events at times 10, 14, 17, 25 and 59 forms the input data for OLE. From this, OLE produces a point estimate of the extinction date (for the example given above, this would be at $t = 143$) as well as an upper and lower 1-alpha confidence interval [in this case (99, 1100) with a 95% CI]. Because the time between sighting events is crucial, OLE also assumes that search effort never drops to zero, and so gaps in sighting record are not simply a product of irregular search forays, a situation that clearly occurs when remote or hostile habitats are rarely searched [e.g. the infrequent search for the Alaotra Grebe, *Tachybaptus rufolavatus*, in 1989, 2004 and 2009, when the species was declared extinct (BirdLife International 2010)]. Although OLE makes no other assumptions, change in the 'observability' of individuals of a species as the population becomes smaller caused by, for example, a change in behaviour could have significant impacts on the accuracy of predictions. On top of this, intrinsic factors (e.g. the life history of a species) or outside pressures (e.g. overhunting) could alter the rate at which a species falls to extinction. Further details are provided in Roberts & Solow (2003) and Solow (2005).

Here we utilize some of the features of microcosm systems to generate long runs of abundance data, preceding eventual extinction, in replicate populations. Varying environmental conditions and community compositions allows us to produce variation in population decline and extinction date. Knowing both the abundances of the populations through time and the size of the habitat in which they live, we generate search regimes to produce data sets of sighting events that vary both in frequency and in temporal spacing. We then examine the effect on the accuracy of extinction estimates, made using OLE, of both observer-controlled (e.g. search effort) and intrinsic (e.g. species identity) factors.

Materials and methods

Fifteen different communities were constructed using every possible composition of four bactiverous protist ciliate species (*Paramecium caudatum*, *Loxocephallus sp.*, *Colpidium striatum* and *Blepharisma japonicum*) (i.e. four single-species communities, six-two-species communities, four-three-species communities and one-four-species community). Five replicates of each community were kept at 15 °C for the duration of the experiment and five replicates at 20 °C, giving a total of 150 microcosms. The microcosms were housed in two incubators, each containing 75 microcosms with one replicate of each community on each of five shelves and the position on a shelf randomized. Shelves were rotated vertically within the incubators every sampling day, and the microcosms were swapped between incubators (but kept at a

constant temperature) every 7 days to minimize any possible incubator effect. Microcosms consisted of lidded plastic petri dishes (ø 100 mm, height 20 mm) containing 50 mL of medium. The medium solution contained 0·1 g of crushed protozoa pellets (Carolina Biological Supply, USA) per litre of Chalkley's medium (Thompson, Rhodes & Pettman 1988). The medium was inoculated with the bacteria *Bacillus cereus*, *B. subtilis* and *Serratia marcescens* on day 2. Communities were established on day 0 by adding a volume of medium containing *c.* 100 individuals of each species from high-density stock cultures. Extinctions were driven by competition for limited resources, and so were due to competitive exclusion or starvation. Evaporative loss over time was replaced with distilled water to avoid any increase in the concentration of salts within the microcosms.

SAMPLING

Sampling was undertaken three times a week for 163 days, at which point the rate of species loss had dropped to only 1–2 extinctions per week and so, given that a high proportion (86%) of *Loxocephallus* and *P. caudatum* populations had gone extinct, the experiment was stopped. To assess presence or absence of a species, each microcosm was placed under a stereoscopic microscope and surveyed for each species at 7·5–30× magnification. If a species was not seen within a 5-min search period, it was assumed to be extinct. In this event, the species was then explicitly searched for during the next sampling period to confirm this. No species that had been recorded as absent in two consecutive samples was observed again.

Sampling to estimate population sizes was based on the methods of Lawler & Morin (1993) and closely mirrored that of Worsfold, Warren & Petchey (2009). Microcosms were mixed and a sample of nine drops was then transferred onto a sterile petri dish, weighed, and the number of individuals of each species counted. The medium was then returned to the microcosm. For rare species that were not present in the first sample, the process was repeated whilst retaining the original sample until a count was obtained. As numbers became very low, individuals were counted within the entire microcosm and the mass of the medium checked with a balance. Simulations were run on the two species (*P. caudatum* and *Loxocephallus* sp.) where multiple extinctions occurred in two, three and four species communities and allowed comparison of both the effect of species richness on predictive accuracy and any interactions between species richness and other variables.

SIMULATING SEARCH EFFORT

Replicates of *P. caudatum* and *Loxocephallus* sp. in which extinction was observed over the experimental time period were treated individually, and to each of these records of population abundance through time four different search regimes of search effort (% of the habitat searched) were applied (see Supplementary Information (henceforth SI) Fig. S1). These regimes were based in part on the search regimes implemented by Rivadeneira, Hunt & Roy (2009): (i) 'constant' effort, (ii) 'decreasing' effort, (iii) 'increasing' effort and (iv) 'realistic' search effort. 'Decreasing' search regime was produced by randomly assigning a search effort (between 0 and 0·95) to the first sampling day and then for each subsequent day search effort decreased linearly (to 0) with the time the species took to go extinct. The 'increasing' search

regime was produced exactly as decreasing, but with a linear increase in the search effort up to 0·95 of the habitat. Thus, both 'increasing' and 'decreasing' search regimes covered a variety of rates of change in the percentage of the habitat searched. As suggested by Rivadeneira, Hunt & Roy (2009), a 'constant' search regime, where the percentage of the habitat searched does not change through time, is perhaps the most unrealistic of the four scenarios and should be viewed as a best-case scenario. Both 'decreasing' and 'increasing' search regimes seem plausible as search effort could be either be increased to see whether a species has gone extinct, or decreased as a species becomes so rare it is assumed to be extinct.

To produce a 'realistic' search effort, change in the area searched through time from a real-world scenario is necessary. The Continuous Plankton Recorder project [Sir Alister Hardy Foundation for Ocean Science (SAHFOS)] provided this information for plankton sampling along known shipping routes. Using a single data set from route A between the years 1958 and 2009 and then calculating the percentage of the vector of route A sampled per month generated a set of search efforts. So, given the average length of route A in 2009 was 173 nautical miles, and each drag (sample event) was for a distance of 10 nautical miles, the search effort in any given month was calculated (with the assumption of no overlap of sampling events) as the number of drags, multiplied by the length of the drag in nautical miles (10) divided by the average length of the route (173 nautical miles). These monthly search efforts were then treated as the search effort for each sample day. The area covered by sampling generated in this way was high (typically around 50%) compared to potential search effort for organisms with large habitat areas. A search of the literature provided no accurate estimates of search effort through time, whilst assessing the status of rare species and so the search regime generated from the SAHFOS data, whilst not perfect, provides at least some example of a real-world sampling regime.

To generate data sets of sighting events under each sampling regime, the original (i.e. complete) data from the microcosms were 'resampled' as follows. Data were converted to total population size within a microcosm. The probability (P) of observing at least one individual at each search event was calculated as:

$$P = 1 - (1 - e)^d$$

Where e is the search effort (fraction of the habitat searched) and d is the abundance of the species within the microcosm (both searched and unsearched). This assumes a random distribution of individuals within a habitat, and thus an increasing chance of observation with increasing abundance and search effort. The 'observation' of a species was then generated as a random event with a probability of success, P. This produced in effect a series of 'historic' sighting events, when extinction is known to have occurred.

Using these sampling regimes produced a series of sighting events at regular intervals, as sighting events could be as frequent as the recorded abundances. This regularity of search effort is probably unrealistic, as highlighted by sporadic searching both temporally and spatially to assess the extinction of *Lipotes vexillifer* (Turvey *et al.* 2007), the Yangtze River dolphin, and again represents a 'best-case scenario'. A more realistic irregular sampling was implemented by running a second simulation of search regimes, identical to the first, but each sighting event produced

was given a 50% probability of being used within the OLE calculation.

Simulations were run between 40 and 95 times, thus each population that suffered an extinction event had between 40 and 95 vectors of sighting events produced. Because there was an element of chance in the sighting events, and so in the number of times point estimates of extinction by OLE occurred, a random sample of 500 from each of the search regimes was used for comparison.

From Solow (2005), optimal linear estimation takes k sighting events and estimates a time of extinction, $\hat{T}_E$, using the form:

$$\hat{T}_E = \sum_{i=1}^{k} w_i t_{n-i+1}$$

Where the weight vector w, length k, is given by:

$$w = (e\prime\Lambda^{-1}e)^{-1}\Lambda^{-1}e$$

e being a vector of k 1's and Λ is a symmetric k by k matrix with typical element:

$$\Lambda_{ij} = \frac{\Gamma(2\hat{v}+i)\Gamma(\hat{v}+j)}{\Gamma(\hat{v}+i)\Gamma(j)} j \leq i$$

where Γ is the gamma function and $\hat{v}$ is an estimate of the shape parameter of the Weibull extreme value distribution given by

$$\hat{v} = \frac{1}{k-1} \sum_{i-1}^{k-2} \log \frac{t_n - t_{n-k+1}}{t_n - t_{i+1}}$$

t_n being the n times a species is observed over the period of time t. So under the assumption, a species is extinct, the upper bound of an approximate $1-\alpha$ confidence interval for $\hat{T}_E$ is

$$T_E^u = \frac{t_n - c(\alpha)t_{n-k+1}}{1 - c(\alpha)}$$

Where

$$c(\alpha) = \left(\frac{k}{-\log\alpha}\right)^{-\hat{v}}$$

Because OLE assumes an extreme Weibull distribution of sightings, it should theoretically only be used with the most recent sighting events (Solow 2005). However, there remains some uncertainty as to the optimum number of sighting events as Collen, Purvis & Mace (2010) found that increasing the number of sightings used (tested to a maximum of 18) increases the accuracy of prediction. Consequently, all four search regimes had extinction estimates produced with the five most temporally recent sighting events ($k = 5$) from a vector of n sighting events, and the total number of sighting events produced by each search regime simulation ($k = n$).

Optimal linear estimation produces a point estimate of the time of extinction, as well as an upper and lower confidence interval. Previously, accuracy of a prediction has been defined as the true date of extinction falling within the 95% confidence intervals

(Rivadeneira, Hunt & Roy 2009). However, given that upper confidence intervals can be thousands of years into the future (Collen, Purvis & Mace 2010), this seems unhelpful. Here we regard accuracy as the proximity of the estimated date of extinction to the actual date of extinction, regardless of the upper or lower confidence interval.

All simulations and analyses were carried out using the R package (R Development Core Team 2011). Statistical tests of the difference between treatments were not employed as the numbers of predictions of extinction date was an arbitrary product of both the number of times simulations were run and the stochastic nature of sighting events.

Results

OBSERVER EFFECTS

Regular sampling provided accurate and precise predictions of time to extinction across all search regimes (Fig. 1). 'Increasing' search effort had the least error (defined as the difference between the actual and predicted date of extinction) associated with predictions (a mean absolute error (overestimation and underestimation combined) of 2·9 days) and the greatest precision ($r^2 = 0.99$, $P < 0.001$ for both $k = 5$ and $k = n$ sightings). 'Decreasing' and 'constant' showed relatively large variation in predictive precision (Fig. 1), although they were still highly correlated (r^2 values higher than 0·97, $P < 0.001$). Using $k = n$ minimized mean error under 'constant', 'increasing' and 'realistic' search regimes. Under a 'decreasing' search regime, mean error was minimized using $k = 5$ sightings, although this decreased the precision of predictions ($r^2 = 0.98$ with $k = 5$ and 0·99 with $k = n$ sighting events, $P < 0.001$ for both). The difference in predictive error across all four search regimes between $k = 5$ and $k = n$ sightings was < 1 day.

Irregular sampling also provided generally accurate predictions under all but a 'decreasing' regime (Fig. 2); however, the precision of predictions was lower compared to regular sampling. The precision of predictions was again higher using $k = n$ for 'constant', 'increasing' and 'realistic' search regimes (Fig. 2) (r^2 values higher than 0·96 and $P < 0.001$ for all three regimes). Both 'increasing' and 'realistic' regimes produced estimates of extinction that were prone to overestimation, a pattern that is present with regular sampling although the magnitude of error is small (Fig. 1). Irregular sampling with a 'decreasing' search regime produced the greatest error. Predictions of extinction time with a 'decreasing' regime dropped from being generally accurate with regular sampling (mean absolute error of 4·1 days when $k = 5$) to providing very poor predictions (mean absolute error of 50·2 days when $k = 5$). This error was almost exclusively accounted for by underestimation of the extinction date, and this pattern was more pronounced as the time to extinction increased (Fig. 2). Precision of estimates with 'decreasing' search effort was also affected by irregular searching ($r^2 < 0.17$, $P < 0.001$ for both $k = 5$ and $k = n$).
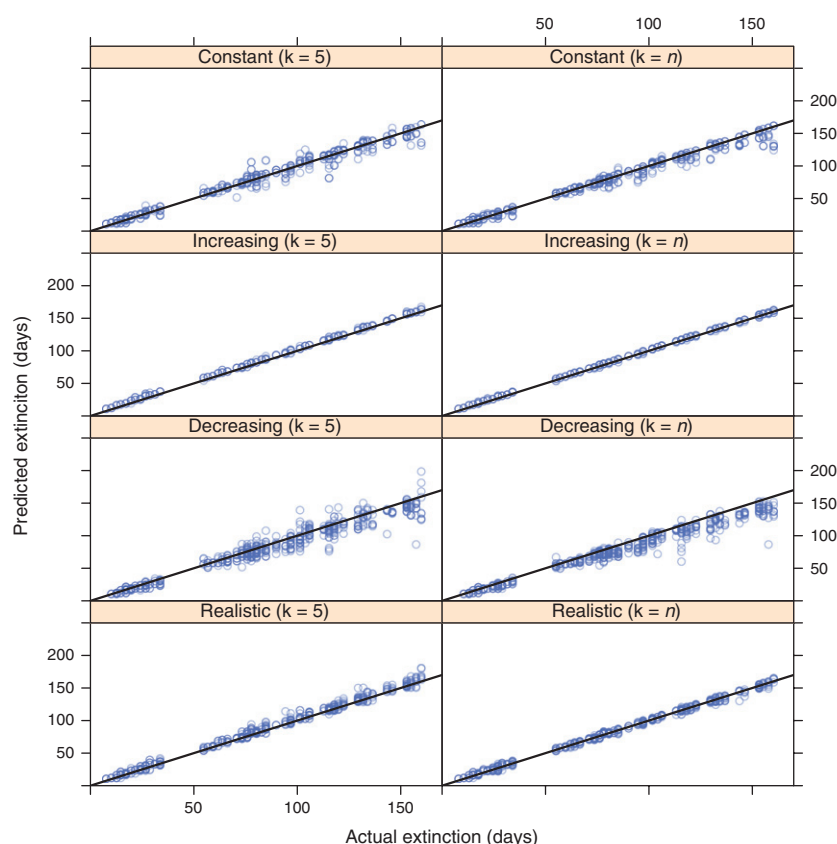

Fig. 1. Plots are of 500 randomly selected predictions of extinction for each species (*P. caudatum* and *Loxocephallus*) where sampling was 'regular', that is, sighting events could potentially occur at every experimental sampling date. Line shows 1 : 1 (or perfect) prediction. Extinction predictions were run with both $k = 5$ and $k = n$ sighting events. Using all available, sighting events produced more accurate and precise results under 'constant', 'increasing' and 'realistic' search regimes. Optimal linear estimation (OLE) produced more accurate predictions when $k = 5$ under a decreasing search regime, although precision was reduced. Gaps in the plots are both a product of the stochastic nature of sampling and a lack of extinctions between days 34 and 55.
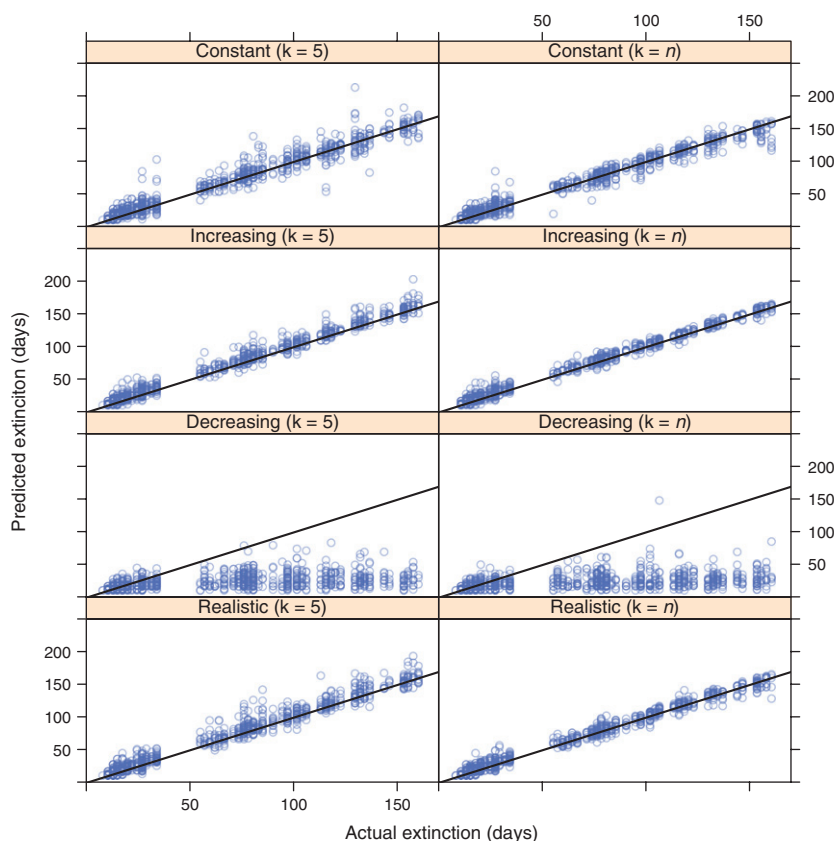
Fig. 2. Plots are of 500 randomly selected predictions of extinction for each species (*P. caudatum* and *Loxocephallus*) where sampling was 'irregular', that is, all sighting events generated by simulations had a 50% chance of being used in the optimal linear estimation (OLE) calculation. So when search effort or population densities are high, the vector of sighting events produced is not a simply a regular sighting every 2–3 days, but has some stochasticity to mimic probable real-world sampling scenarios. Line shows 1:1 (or perfect) prediction. Extinction predictions were run with both $k = 5$ and $k = n$. Irregular searching greatly affect the precision of OLE. 'Decreasing' search effort showed a significant loss of both precision and accuracy, with a large increase in underestimation of extinction date. Gaps in the plots are both a product of the stochastic nature of sampling and a lack of extinctions between days 34 and 55.

The percentage of the habitat searched (held constant through time and with regular sampling) affected the accuracy of predictions, with area of the habitat searched positively correlating with precision (Fig. 3a,b). The majority of this error is underestimation (Fig. 3a). Overestimation accounts for only a small proportion of the magnitude of the total error, and above 7% search effort there is remarkably little variation in error. Increasing the percentage of the habitat searched from 1% to 4% produced a large decrease in absolute error (an average of 3·9 days; Fig. 3b). Similarly an increase from 4% to 10% continued to increase precision, albeit at a lesser rate. Increasing search effort from 10% to 95% produced an increase in precision of only 2·1 days.

The number of sighting events (k) used in the OLE calculation showed no clear relationship to accuracy (Fig. 4). 'Increasing' and 'realistic' search regimes showed little variation in error as the number of sightings used
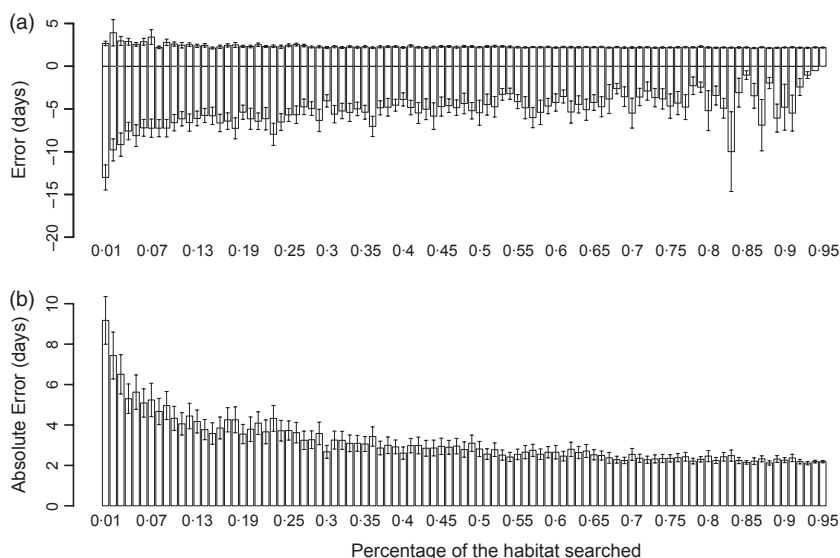

Fig. 3. All error bars are one SE, missing SE bars indicate means comprised of < 3 values. Data are for 'constant' search regime with regular searching. a) The majority of error in prediction is accounted for by underestimation of the date of extinction. Overestimation shows remarkably little variation as search effort increase above 7%. b) Absolute error shows the increase in overall predictive accuracy as search effort is increased. Above 10% of the habitat sampled, there appears to be relatively little increase in predictive precision.
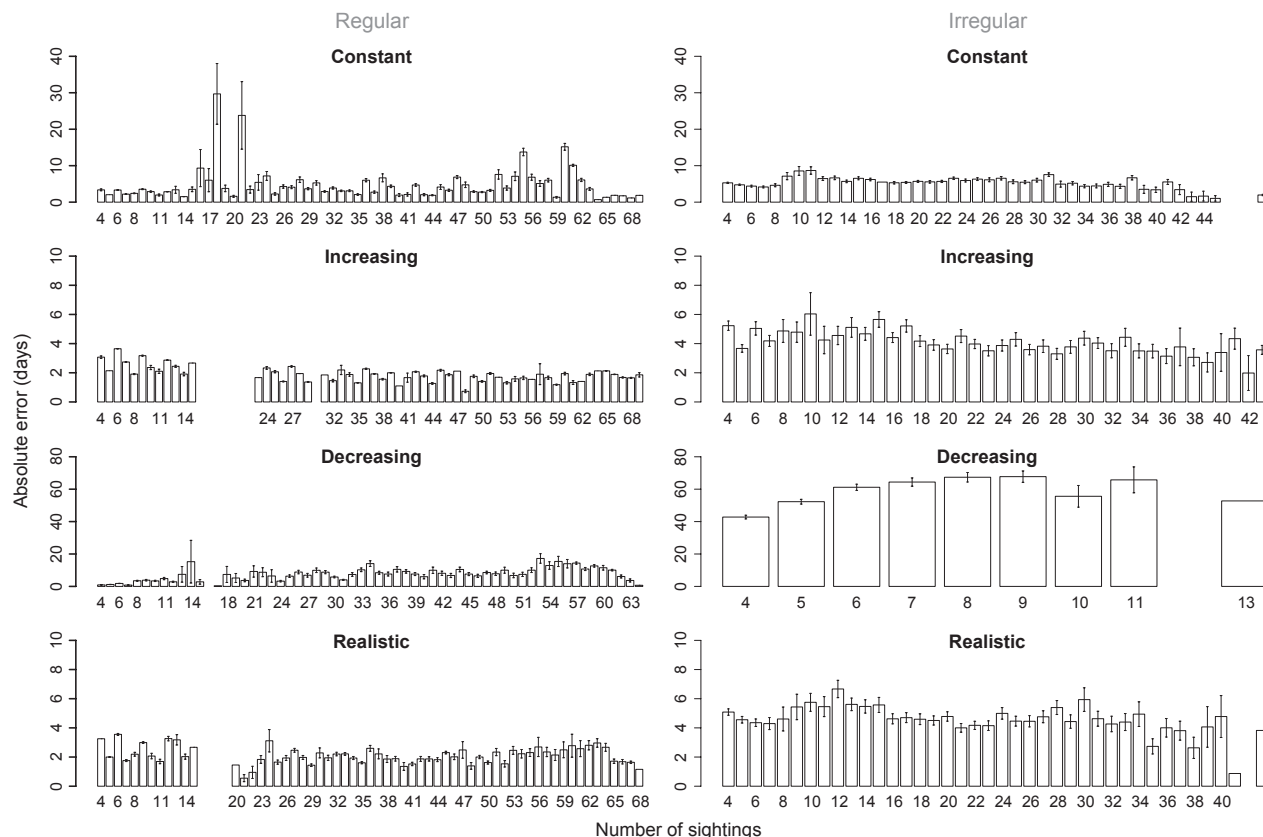
Fig. 4. Error bars are 1 SE, missing SE bars indicate means comprised of < 3 values. The number of sighting events used to calculated the extinction date of a species using optimal linear estimation (OLE) altered the mean error associated with predictions of time to extinction. 'Decreasing' search regime with irregular sampling appeared to have a slight positive correlation between number of sightings used and error of predictions. This supports the finding that using the last five sighting events only for 'decreasing' search regime produced on average less error of predictions. The other three search regimes appear to have no clear relationships between the number of sightings used and the error associated with the predictions. Gaps in the plots are both a product of the stochastic nature of sampling and a lack of extinctions between days 34 and 55.

increased. 'Constant' search regime peaked in error when $k = 18$, although this can probably be attributed to a low number of estimations of extinction made using 18 sightings ($n = 16$, mean $= 231$), a product of both the stochastic nature of sighting events and a lack of extinctions between days 34 and 53. 'Decreasing' search regime with irregular sampling generated around an order of magnitude more error than regular sampling. With a 'decreasing' regime and regular sampling, the number of sightings used shows a weak positive correlation to error, a pattern that becomes more obvious when sampling is irregular.

INTRINSIC EFFECTS

In this experiment, systems differed in three ways: the number of species present, the identities of those species, and the temperature at which they were kept. These three variables all affected the rate at which species were driven to extinction (Supplementary Information Fig.S2). However, predictions of extinction produced using OLE appear in general to be mostly unaffected by these variables, with a few notable exceptions (Fig. 5). As temperature increases from 15 °C to 20 °C mean error in general

decreases, but the majority of this error decreases by an insignificant amount (1–2 days) with a few cases having a much larger reduction in error. Increased species richness has almost no effect on error. The identity of the species is associated with large variation in the error of the predicted date of extinction, with *P. caudatum* almost always having greater error, and in a few cases, this error is orders of magnitude larger than found in comparative *Loxocephallus* treatments.

Discussion

Given the problematic nature and history of determining whether a species is extant or extinct (Fisher & Blomberg 2011), and the consequences of getting it wrong (Collar 1998; Jackson 2006), the ability to accurately infer extinction status would be an invaluable tool for conservation. For many species, sighting data are all that is available, so if sighting-based predictors are accurate then this provides a tool of great utility. For those species where sighting data exist, OLE appears, in general, to provide accurate and precise predictions of time to extinction. OLE has, unsurprisingly, proved sensitive to factors both
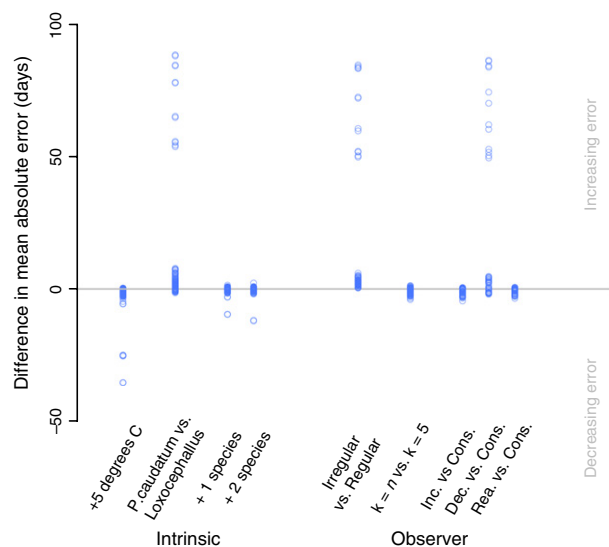
Fig. 5. Change in error associated with a change in each variable was calculated by subtracting the range of mean errors, grouped by similar intrinsic and observer attributes, from one another. Thus, the change in error due to an increase in degrees is the mean errors associated with the 20 °C treatment predictions minus the mean errors associated with the 15 °C treatment predictions, with all other variables kept constant. Thus, all differences in mean errors are relative to the reference point, not to the actual date of extinction. Cons. = 'Constant' search effort, Inc. = 'Increasing' search effort, Dec. = 'Decreasing' search effort, Rea. = 'Realistic' search effort.

inherent to the system it is applied to, and those arising through the observation process. Under ideal conditions (and assuming an inability to search 100% of a species' habitat) sighting data collected using the highest possible search effort held constant through time with regular, frequent sampling would yield the most accurate and precise predictions of a species' extinction date.

However, the major advantage of OLE, and extinction estimators in general, is in its ability to infer extinction under data depauperate conditions. Irregular searching and low search effort can produce large variations in both the accuracy and precision of estimates of extinction. At low search efforts, small increases in the search area provide large decreases in error (Fig. 3a,b). Search effort correlates positively with the potential to overestimate an extinction event (Fig. 3a), and this correlation between search effort and overestimation explains the trends seen in both 'increasing' and 'decreasing' search effort (Figs 1 and 2) and makes intuitive sense as searching large proportions of a habitat prior to an extinction event means the chances of observing an individual increase, and consequently shift the predicted extinction date forward through time (and visa versa). Overestimation of the extinction date was more common than underestimation (a ratio of *c.* 3 : 1), although this is almost certainly a product of low search efforts simply not producing enough sighting events to use in OLE, thus more estimations were made at higher search efforts. That being said

where underestimation of the extinction date was present it was generally of greater magnitude than overestimation (means of 6·4 and 3·9 days, respectively).

Our results suggest that change in search effort through time, either through increasing or decreasing resource availability or fluctuations brought on by opportunistic sampling, can reduce predictive accuracy, especially if searching is irregular [where patterns observed under a regular sampling appear magnified (Figs 1 and 2)]. Unfortunately, in many situations, it may not be apparent whether the search effort has changed (for example, where there may be habitat loss, poor records of the area sampled or areas wrongly identified as suitable habitat). The interaction between changing search effort, search regularity and species identity is one that should be considered carefully, especially if applying OLE or other extinction indices to species whose life histories are particularly long or unknown.

The estimations of extinction with exceptionally high associated error (Fig. 5) all share some common features: they are predictions of extinction of the species *P. caudatum* at 15 °C under a decreasing search effort with an irregular search pattern. The population dynamics of *P. caudatum* at 15 °C (Supplementary Information Fig. S2) are such that for relatively long periods of time population abundances are very low prior to an extinction event: a population of *P. caudatum* survived for 23 days with fewer than five individuals, and for 9 days with a single individual. With these dynamics, a 'decreasing' search regime and, importantly, irregular searching, the chances of missing the final few individuals for long periods of time becomes almost a certainty. Where populations fell rapidly to extinction [e.g. *Loxocephallus* in species rich communities at 20 °C (Supplementary Information Fig. S2)] predictions produced far less error, even with irregular searching and a 'decreasing' search regime. This supports work by Rivadeneira, Hunt & Roy (2009) who found that model populations with rapid population decline to extinction produced significantly better predictions of time to extinction than those with slower rates of decline, across a range of extinction predictors. It should be noted, however, that because OLE requires a minimum of five sighting events to make a prediction, a larger proportion of the population decline is covered by the same number of sighting events for species with rapid rates extinction. This explains some of the overestimation of extinction date seen with a 'decreasing' search regime and irregular searching where extinctions occurred early in the experiment (Fig. 2).

There has been some uncertainty as to the optimum number of sightings to use in the OLE calculation (Solow 2005; Collen, Purvis & Mace 2010). Because OLE assumes a Weibull extreme value distribution, large numbers of sightings may violate the asymptotic argument on which the model is based (Solow 2005). Rivadeneira, Hunt & Roy (2009) identified during preliminary analysis that including large numbers of sightings increased the upper

bounds of the estimates. However, when considering the estimated date of extinction [as opposed to upper confidence intervals as in Rivadeneira, Hunt & Roy (2009)], this was found not to be the case as all regimes other than 'decreasing' showed overall a greater precision when an unlimited number of sightings were used (Figs 1 and 2), and there appeared to be no obvious relationship between error and the number of sighting events (Fig. 4). Across all search regimes (and under regular or irregular sampling), the difference in mean absolute error between using either $k = 5$ or $k = n$ was never > 3 days. Consequently, the use of all available sightings ($k = n$) for real-world calculations seems reasonable as for the majority of search scenarios this improved both the precision and accuracy of predictions.

Although on average OLE produces accurate estimations of time to extinction, within this there is both over-estimation and underestimation of the extinction date, and when search effort is low and sampling irregular, care should be taken when declaring a species extinct. Previously the confidence intervals (typically set at 95%) produced by OLE and other extinction indices have been used to reduce the chance of erroneously declaring a species extinct (Rivadeneira, Hunt & Roy 2009). However, this overcautious approach serves to increase uncertainty, as confidence intervals are typically wide, spanning centuries or millennia (Collen, Purvis & Mace 2010). Within this experiment, upper 95% confidence intervals were occasionally over 700 days after the predicted date of extinction (with regular sampling) and over 3000 days after the point estimate of extinction with irregular sampling (Supplementary Information Fig. S3). The use of $k = n$ sighting events reduced the number of extremely high 95% confidence intervals, but only with regular sampling. Within this experiment, 95·3% of observed extinctions were encompassed by the upper and lower 95% confidence intervals when sampling was regular, and 82·4% when sampling was irregular. However, an alternative approach is to employ a fixed time period, or 'safety-net', after the point estimate of extinction produced by OLE before declaring a species as extinct. This period should be a balance between the risk of investing time and money on trying to save a species that is extinct and removing funding early and losing a species that could be saved, whilst reducing the uncertainty produce by wide confidence intervals. This is a potentially novel way of applying OLE and using this approach, with a 'safety-net' period of 5 days added to point estimates of extinction, 90% of extinction events were encompassed with regular sampling and 73% with irregular. When this 'safety-net' is increased to 10 days, these figures grow to 95% and 81% respectively.

Estimates of minimum generation times for *P. caudatum* and *Loxocephallus* [between 0·73 and 0·84 days (*SI*)] suggest that it is both convenient and appropriate to consider a day within the microcosm system to be roughly analogous to a year in a natural system. These estimates would then be roughly equivalent (0·73–0·84 years) to, say, time of first reproduction of many mammal species (Wootton 1987). If a day can be viewed as roughly analogous to a year (and the species in question has generation times roughly equivalent to those presented here), then imposing a 'safety–net' of 5–10 years, depending on the regularity of historical searching, after the predicted extinction date before a species is declared extinct would reduce any potential 'Romeo Error' (declaring a species as extinct when it is still extant) (Collar 1998). However, this is a rough estimate, not a prescriptive value. Species with significantly different life histories, fecundity, ability to recover from small populations and a myriad other factors, could influence the time period required as a 'safety-net'.

Conclusions

Optimal linear estimation has proved sensitive to both intrinsic and observer-controlled variables; however, on average, it appears to accurately predict the date of an extinction event. Use of the inferred date of extinction generated by OLE, as apposed to the confidence intervals considered previously, provides more sensible estimates of extinction that are more widely applicable to real-world scenarios.

The accuracy of predictions was highly dependent on search effort, search regularity and change in search effort through time. Increasing search effort from 1% to 10% of a species' habitat rapidly reduces the error associated with predictions, whilst searching more than 10% produces a negligible increase in predictive precision. Decreasing search effort through time and irregular searching are liable to produce the greatest predictive error.

Intrinsic properties of the system, such as populations that subsist for extended periods of time at very low densities, can magnify the error, specifically underestimation of extinction date. When low search effort, irregular searching and species identity interact error can be orders of magnitude larger than otherwise found. Extensive testing in this model system suggests that to minimize error in real-world scenarios using all possible sighting events will eliminate most of the error associated with differing search regimes, and implementing a 5 to 10-year 'safety-net' before declaring a species as extinct will minimize any 'Romeo Error'. Observer controlled and intrinsic factors can significantly affect predictive accuracy and should be taken into account when testing extinction predictors in the future.

Acknowledgements

References

Barnosky, A.D., Matzke, N., Tomiya, S., Wogan, G.O.U., Swartz, B., Quental, T.B., Marshall, C., McGuire, J.L., Lindsey, E.L., Maguire, K. C., Mersey, B. & Ferrer, E.A. (2011) Has the Earth's sixth mass extinction already arrived? *Nature*, **471**, 51–57.

BirdLife International. (2010) Tachybaptus rufolavatus. IUCN 2011. IUCN Red List of Threatened Species. Version 2011.2.

Collar, N.J. (1998) Extinction by assumption; or, the Romeo Error on Cebu. *Oryx*, **32**, 239–244.

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.

Elphick, C.S., Roberts, D.L. & Reed, M.J. (2010) Estimated dates of recent extinctions for North American and Hawaiian birds. *Biological Conservation*, **143**, 617–624.

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.

Fisher, D.O. & Blomberg, S.P. (2012) Inferring Extinction of Mammals from Sighting Records, Threats, and Biological Traits. *Conservation Biology*, **26**, 27–67.

Fitzpatrick, J.W. (2006) Response to Comment on "Ivory-billed Woodpecker (Campephilus principalis) Persists in Continental North America". *Science*, **311**, 1555b.

Fitzpatrick, J.W., Lammertink, M., Luneau, M.D., Gallagher, T.W., Harrison, B.R., Sparling, G.M., Rosenberg, K.V., Rohrbaugh, R.W., Swarthout, E.C.H., Wrege, P.H., Swarthout, S.B., Dantzker, M.S., Charif, R.A., Barksdale, T.R., Remsen, J.V., Simon, S.D. & Zollner, D. (2005) Ivory-billed woodpecker (Campephilus principalis) persists in continental North America. *Science (New York, N.Y.)*, **308**, 1460–1462.

Gotelli, N.J., Chao, A., Colwell, R.K., Hwang, W.-H. & Graves, G.R. (2011) Specimen-Based Modeling, Stopping Rules, and the Extinction of the Ivory-Billed Woodpecker. *Conservation Biology*, **26**, 47–56.

Jackson, J.A. (2006) Ivory-billed woodpecker (Campephilus principalis): hope, and the interfaces of science, conservation, and politics. *The Auk*, **123**, 1–15.

Keith, D.A. & Burgman, M.A. (2004) The Lazarus effect: can the dynamics of extinct species lists tell us anything about the status of biodiversity? *Biological Conservation*, **117**, 41–48.

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

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

McInerny, G.J., Roberts, D.L., Davy, A.J. & Cribb, P.J. (2006) Significance of Sighting Rate in Inferring Extinction and Threat. *Conservation Biology*, **20**, 562–567.

Memmott, J., Waser, N.M. & Price, M.V. (2004) Tolerance of pollination networks to species extinctions. *Proceedings of the Royal Society B Biological Sciences*, **271**, 2605–2611.

R Development Core Team. (2011) *R: A Language and Environment for Statistical Computing*. R Development Core Team, Vienna, Austria.

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.

Rös, M. & Pineda, E. (2009) Apparent extinction or insufficient sampling?: comment on "Deforestation and apparent extinctions of endemic forest beetles in Madagascar". *Biology Letters*, **5**, 651–652.

Rout, T.M., Heinze, D. & McCarthy, M.A. (2010) Optimal allocation of conservation resources to species that may be extinct. *Conservation Biology*, **24**, 1111–1118.

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

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

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

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

Thompson, A., Rhodes, J. & Pettman, I. (1988) Culture collection of algae and protozoa: catalogue of strains. *Natural Environment Research Council*, **5**, 164.

Turvey, S.T., Pitman, R.L., Taylor, B.L., Barlow, J., Akamatsu, T., Barrett, L.A., Zhao, X., Reeves, R.R., Stewart, B.S., Wang, K., Wei, Z., Zhang, X., Pusser, L.T., Richlen, M., Brandon, J.R. & Wang, D. (2007) First human-caused extinction of a cetacean species? *Biology Letters*, **3**, 537–540.

Wootton, J.T. (1987) The Effects of Body Mass, Phylogeny, Habitat, and Trophic Level on Mammalian Age at First Reproduction. *Evolution*, **41**, 732–749.

Worsfold, N.T., Warren, P.H. & Petchey, O.L. (2009) Context-dependent effects of predator removal from experimental microcosm communities. *Oikos*, **118**, 1319–1326.

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Data S1. Generation times.

Fig. S1. Four search regimes were simulated, with the percentage of the habitat searched (search effort) either being held constant, or changing through time according to the regime.

Fig. S2. Temperature and species richness both negatively correlated with time to extinction, whilst species identity changed the time to extinction by up to an order of magnitude.

Fig. S3. Upper and lower 95% confidence intervals (CIs) are often wide, and are dependent on both the number of sighting events used in the OLE calculation and the search regime.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.
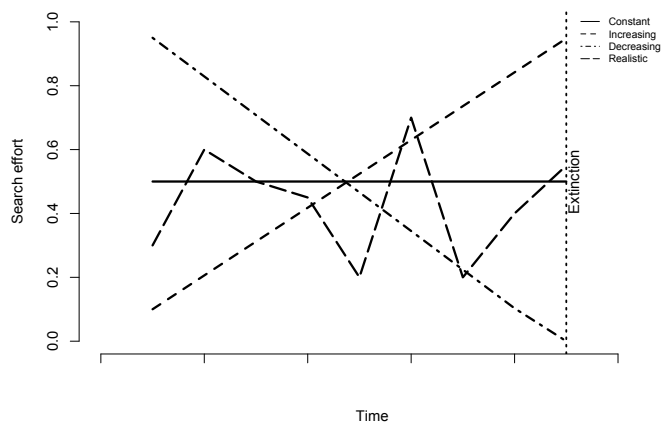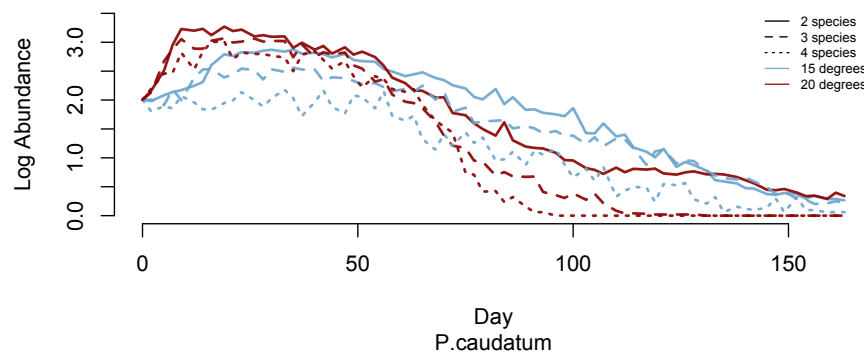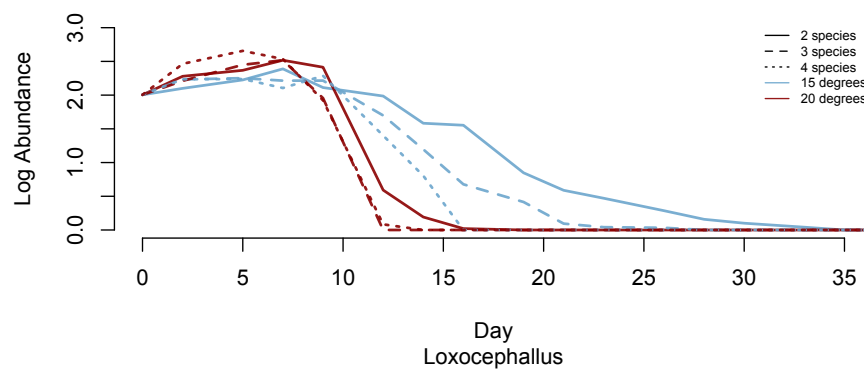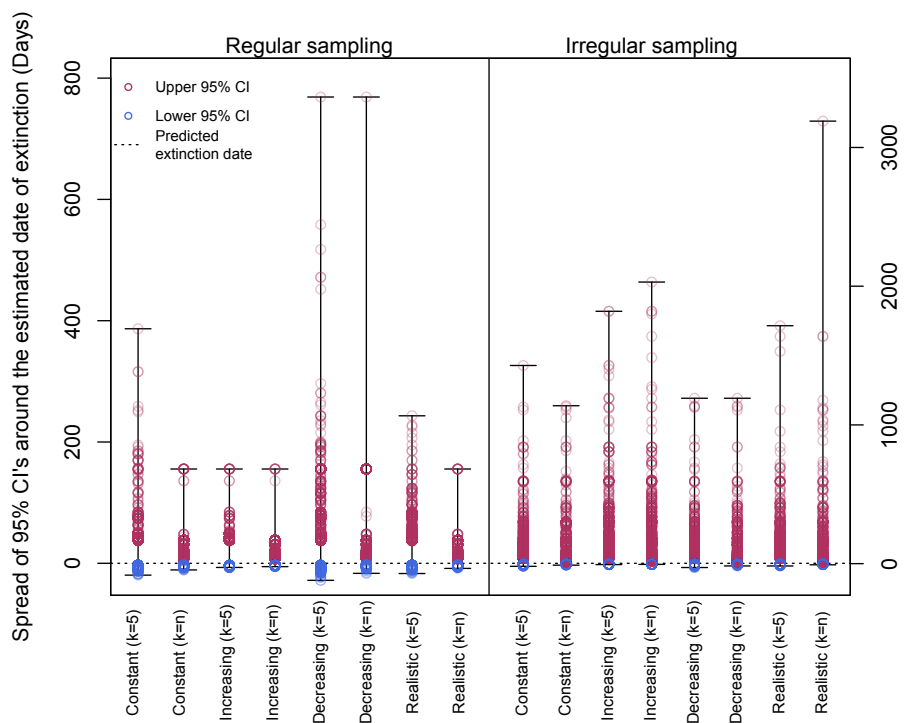
Figure 1.

Figure 2.

Figure 3.