A fast re-sampling method for using reliability ratings of sightings with extinction-date estimators

BARRY W. BROOK[1,2*], JESSIE C. BUETTEL[1,2], IVAN JARIĆ[3,4]

[1] *School of Natural Sciences, University of Tasmania, TAS 7001, Australia*

[2] *ARC Centre of Excellence for Australian Biodiversity and Heritage, University of Tasmania*

[3] *Biology Centre of the Czech Academy of Sciences, Institute of Hydrobiology, České Budějovice, Czech Republic*

[4] *Department of Ecosystem Biology, Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic*

[*] Corresponding author: barry.brook@utas.edu.au

Abstract

The pattern of sightings of a species that is rare, and then no longer observed, can be used to estimate its extinction date. However, other than physical captures or specimens, the veracity of any sighting is ambiguous, and should be treated probabilistically when used to infer extinction dates. We present a simple yet powerful computational approach for incorporating observational reliability into extinction date estimators (EDE). Our method: (i) combines repeated within-year sightings probabilistically, (ii) samples observations using the reliability score as an inclusion probability, (iii) infers a probability distribution and summary statistics of extinction dates with any EDE, and (iv) computes the frequency distribution of the extinction date. We applied this method to eight exemplar sighting records covering a range of lengths, sighting rates and uncertainties, using a variety of statistical EDEs, and compared these results with a threshold approach for selecting sightings. We also demonstrated a robust coverage of 'true' extinction dates based on selected real-world examples of rediscovered species and confirmed extinctions, and simulated sighting records. Our approach represents a powerful generalization of past work because it is not predicated on any specific method for inferring extinction dates, and yet is simple to implement (with R script provided).

***Key words**: extinction model, extinction time, sighting record, uncertainty, computational ecology, species persistence*

INTRODUCTION

The occurrence and timing of extinctions are notoriously difficult to confirm. Extreme rarity typically precedes extinction (leading to detection problems and the need for substantial search effort), and the event is, by definition, an absence of all individuals. Consequently, extinction is usually inferred via proxy information, such as records of observations. These ‘sightings’ can come in many forms, including museum/herbaria specimens, live-animal captures, acoustic recordings, photographs/film, hair/faecal samples, footprints, and visual observations. Yet, except for verified *post-mortem* specimens or live captures, the veracity of these sighting types is ambiguous, due to possible misidentifications, illusions or deceit. This is problematic, because the statistical inferences made using various model-based pattern analyses of sighting records (known as extinction-date estimators: EDE) can be highly sensitive to the decision to reject or include indirect or otherwise ambiguous observations (Lee et al. 2014). Indeed, the question of how to treat anecdotal or inconclusive data has led to much debate regarding the presence, species range, and population dynamics of a suite of species (e.g., ivory-billed woodpecker, Californian wolverine: McKelvey et al. 2008). However, ambiguous sightings of rare species are often made by trained biologists, experienced amateur naturalists, or well-meaning members of the public, making it hard to argue that they should be disregarded entirely (Lee et al. 2014). Decision errors caused by either the false rejection of valid sightings, or the acceptance of mistaken observations, can have conservation implications and result in economic or opportunity costs.

Existing methods for handling mixed-certainty sighting records either classify observations as unambiguous or ambiguous based on sighting type, or else assign a specific reliability score, which can be thought of as the probability that the sighting is a true observation of the species (Boakes et al. 2015). This process might involve consultation with experts (Elphick et al. 2010) or application of a classification-scoring system to assign ‘plausibility ratings’ based on observation type (e.g., BirdLife International; see Lee et al. 2014). For example, McKelvey et al. (2008) developed a set of evidentiary standards that increases in rigor depending on the species rarity. One approach is to thereafter compile a final sighting record for use in EDEs that is based on accepting or rejecting observations based on sighting type or some threshold ‘inclusion probability’ derived from a scoring system. Jarić and Roberts (2014) proposed an alternative, with all observations being used, but weighted within an EDE based on their sighting reliability; this permits the inclusion of controversial sightings (with lower probabilities of being a true sighting of the species). Most EDE methods for mixed-certainty data are based on a Bayesian framework, whereby observations are grouped (and modelled) as separate ‘certain’ or ‘uncertain’ sighting records (e.g., Solow et al. 2012, Thompson et al. 2013, Lee 2014, Solow and Beet 2014). However, these require prior information on population dynamics that is rarely available, are mathematically or computationally demanding, and can be difficult to compare to more traditional frequentist methods (e.g., Solow 1993, Saltre et al. 2015). This has tended to discourage their widespread use (Boakes et al. 2015); as noted by Lee (2014), the main reason for the widespread use of the Solow (1993) method is its user-friendly implementation, via a single equation. Simplicity of a method also increases its transparency and its accessibility as a decision tool (Rout et al. 2010). The only frequentist method to incorporate sighting reliability (Jarić and Roberts 2014) is simple, but it does not really constitute a new model. However, it remains unclear whether the probabilistic treatment of sighting veracity (or classification of observations into different reliability types) has a general effect of improving or biasing model-based estimates of extinction probability.

Here we present a novel, generalized approach to account for observational reliability using any sighting-based EDE, hereafter referred to as weighted re-sampling (BBJ). The method is easy to implement (and we supply optimized R code [r-project.org] to run the model), can combine repeated within-year sightings, and computes summary statistics and the frequency distribution of extinction times. We demonstrate its application on historical records with different lengths, sighting rates and underlying uncertainties (using a variety of statistical EDEs) and compare our results with established frequentist and Bayesian approaches. We also evaluate the method's performance against real-world examples of re-discovered species, 'confirmed' extinctions, and simulated records.

METHODS

Conceptually, the approach is straightforward. The idea is to repeatedly re-sample without replacement from a sighting record of length *n*, with each observation's sighting reliability being used as a surrogate probability for randomly determining its inclusion in (or rejection from) the sample sighting record. The use of many replicate sample records (~1 000 or more) will yield a stable result in terms of the distribution of re-sampled TE, and its moments. The resulting stochastic sighting-record vectors (v), of length $3 < n_v \leq n$, can then be used as input for any EDE model to estimate a distribution of the time of extinction (TE). Sighting reliabilities can either be specified on a continuous bounded scale (between 0−1) or aggregated in classes (e.g., physical evidence, expert opinion, controversial) that are each assigned an associated reliability. A notable advantage of the BBJ method is that the confidence bounds of any EDE can be determined empirically in this way, with sighting uncertainty handled in the data pre-processing step rather than as an assumption-driven characteristic of the EDE model itself. It can be applied to any EDE that does not otherwise attempt to account explicitly for ambiguous observations (other than via thresholding). An in-depth justification for the re-sampling approach is given in Appendix S1: Section S1.

In the case of datasets with multiple ambiguous sightings per year, the method can be adapted by a simple extension to account for the within-year sightings' individual reliabilities. If p_{ij} is a sighting reliability of i^{th} sighting out of n sightings in year j, then the probability that there is at least one true sighting in that year (p_j) will be:

$$p_j = 1 - \prod_{i=1}^{n} 1 - p_{ij}$$

In the algorithm, years with multiple sightings are thus processed first and the sighting record for that year is replaced with a single sighting of reliability p_j. We note that the same extension can be also applied to the method of Jarić and Roberts (2014) or any other discrete-time method (e.g., Thompson et al. 2013, Lee 2014, Lee et al. 2014). We also propose an additional improvement to the Jarić and Roberts (2014) approach, in which the most-likely year of first and last observation are both estimated, as detailed in Appendix S1: Section S2. The sensitivity of the extinction inference to each sighting within a record can be evaluated using jackknifing, whereby each sighting is omitted, in turn, as the algorithm is re-run.

Our R script for implementing the algorithm (provided as `DataS1.zip`) is function-driven and appropriately commented for ease of interpretability. Output includes the median TE (reported in the same units as the sighting record, typically as calendar years, although any positive integer is possible), the upper 95% confidence bound of the estimate (2.5th percentile of the frequency distribution of the TE values), and cumulative probability of persistence by year. Aside from these two EDEs, other methods that are currently implemented in the R code are Robson and Whitlock (1964), Strauss and Sadler (1989), Solow (1993), hereafter S93, ordinary least squares estimator (OLE) approach of Roberts and Solow (2003), hereafter OLE, McInerny et al. (2006), and the modified version of the Jarić and Roberts (2014) method (discussed above). The R code is modular, allowing new EDE functions to be added, provided they take a sighting record (specified as a vector of years, or positive integers) and return a time of extinction.

To evaluate the sensitivity of our new method to different circumstances, we applied it to eight real-world cases and compared its results with those EDEs which remove sightings that do not meet a reliability threshold (the ‘threshold approach’). In practice, this threshold can be quantitative or qualitative (e.g., based on sighting type). In our test, we used a sighting-type threshold, including only those based on physical evidence (live or dead specimens or parts thereof). The species examples we use represent: (i) four with well-known sighting records (baiji dolphin (*Lipotes vexillifer*), Barbary lion (*Pathera leo*), ivory-billed woodpecker (*Campephilus principalis*) and O‘ahu nukupu‘u (*Hemignathus lucidus*)), containing a mix of unambiguous (reliability = 1) and ambiguous (reliability < 1) records, and which have been used frequently in previous papers that have proposed or evaluated methods for estimating extinction from sighting records (e.g., Jarić and Roberts 2014, Lee et al. 2014, Boakes et al. 2015, Lee et al. 2015); (ii) two (the night parrot (*Pezoporus occidentalis*) and noisy scrub bird (*Atrichornis clamosus*)) with past sightings (of mixed certainty) spread across multiple years, followed by an extended interval (typically of many decades) of apparent absence, after which they were re-discovered (“Lazarus species”); and (iii) two (Bramble Cay melomys (*Melomys rubicola*) and Alaotra grebe (*Tachybaptus rufolavatus*)) that were regularly reported with a mix of sighting uncertainty, and might persist, except that other evidence (e.g., exhaustive searches across their range, or a complete loss of suitable habitat) implies that extinction is ‘definite’.

In general, a robust method should indicate a low probability that a species from the Lazarus group is extinct, and *vice versa* for the definite group. A poor performance in such cases implies, respectively, that the method is either overly liberal or conservative (i.e., prone to either Type I or II errors). Analysis of additional examples, illustrating a diversity of sighting-record characteristics, are also provided in Appendix S1: Section S2 and the complete sighting records are available as a R script in `DataS1.zip`. Data were sourced from the literature cited above, as well as the Global Biodiversity Information Facility (gbif.org), Atlas of Living Australia (ala.org.au), IUCN (iucnredlist.org) and grey literature.

As an adjunct to the real-world case studies, we also ran a sensitivity analysis with simulated data, following the general approach by Rivadeneira et al. (2009) and Jarić and Roberts (2014). To do this, we generated mixed-certainty sighting records stochastically, with a duration of 50–80 years and a fixed per-year sighting probability ranging among simulations from 0.1–0.3; this resulted in 5–24 observations per record and a known (random) time of extinction set to occur at sometime within the second half of the sighting record. Although a simplified characterization of reality, these simulated data allowed us to explore the general circumstances under which the new method performs well (or poorly) compared to other approaches, when there is a known result. A detailed description of this simulation test is provided in Appendix S1: Section S3 and R scripts in `DataS1.zip`. A caveat of the simulation approach is that because we sought to make the BBJ re-sampling method and threshold application of S93 and OLE comparable, we had to use the same 1000 simulated sighting records (one set for constant and one for declining sighting probability) for all methods. Thus, all simulated records had to meet the strictest criterion when setting the threshold of 0.8, to yield >=5 sightings; consequently, all such records tended to be 'good' datasets. For testing lower thresholds such as 0.4 or 0.2, much worse records (i.e., with poorer overall reliability) could be also used, and these 'poor' records are *a priori* likely to be those where the BBJ method ought to be superior to the threshold methods. Overall, the need for selection equivalence will tend to bias the sensitivity results in favor of threshold methods.

RESULTS

Figure 1 illustrates the consequences to the estimation of TE for the observational records of four well-known species, using four alternative EDEs. In these examples, we compared three approaches to associating probabilities with sighting records: (i) use of the mid-point probabilities of BirdLife International (BL-M; where the veracity (inclusion probability) of physical records = 0.85, expert observations = 0.7 and controversial sightings = 0.25), (ii) retention of physical records only (P-only; where physical = 1, expert = 0, controversial = 0), and (iii) a choice to heavily down-weight—but not exclude—non-physical records (Extreme; where physical = 0.99, expert = 0.5 and controversial = 0.01); however, many alternatives are possible and available in the R script in `DataS1.zip`.

For real-world species with a 'known fate', the results detailed in Table 1 illustrate the similarities and differences arising for predictions for Lazarus and definite species, using the S93 and OLE models in both their original form and when sighting uncertainty is included using our new re-sampling approach. In most cases, 'coverage' of the actual event (i.e., persistence for the Lazarus species, and extinction for the definite species) was improved when the re-sampling approach was used to incorporate ambiguous sightings along with unambiguous sightings (physical specimens). For example, for a Lazarus species (night parrot), both models predicted extinction before 1989 (the year prior to rediscovery) when only unambiguous sightings were used. But when ambiguous sightings were included via the BBJ method, both the S93 and OLE predicted that persistence was more likely than extinction. For a definite species (Bramble Cay melomys), the OLE predicted that the species was likely extinct by its declared date of 2014, whereas S93 was somewhat over-optimistic, irrespective of whether ambiguous sightings were included or excluded.

The efficacy of the re-sampling approach is further supported by the results of the simulation study, where the extinction time is known. In general, the performance of S93 and OLE models were strongly influenced by the choice of the reliability threshold value, while the BBJ method performed well without a need to make arbitrary thresholding decisions (Table 2). The BBJ re-sampling showed consistently lower mean absolute differences between the true and estimated time of extinction compared to S93 and OLE with ambiguous sightings excluded, under both constant and declining sighting rates and rejection thresholds. The BBJ method outperforms threshold approaches in terms of bias in all circumstances, and coverage of confidence intervals in most cases, except for when S93 is used in the case of a declining sighting rate over time, when a threshold rejection approach seems to fare better.

DISCUSSION

We have developed a fast, generalized and simple computational method for incorporating sighting reliabilities into any discrete-time extinction date estimator and used selected examples to demonstrate its utility. Our approach does not replace any existing EDE model, but it does mitigate against arbitrary data filtering (by obviating the need to decide upon some reliability threshold for including or excluding sightings). For any situation where an index of reliability can be estimated for species records, this method allows for explicit accounting of sighting uncertainty, either for quality types (e.g., physical specimens, expert opinion, controversial sightings, etc.) or for each observation individually. That said, we recommend that any analysis using sighting reliabilities includes a sensitivity analysis of the reliability scores, such as evaluating lower, upper and best-estimate bounds, to gauge how sensitive the extinction probability and/or date is to probabilities associated with different reliability scores or sighting-quality types. The approach is universal in not being predicated on any specific EDE model or statistical assumptions for inferring extinction dates. It is also ‘future proof’, in that—because the method focuses on probabilistic sampling of the sighting record—it could be used with yet-to-be-developed EDE methods of arbitrary complexity.

Key advantages of our method are its ease of implementation (with modular R code provided) and its ability to undertake a flexible sensitivity analysis of the factors that most contribute to uncertainty in the estimation of extinction times for rare species. Although many EDE methods have been proposed in the literature, few have been available to use 'off the shelf', leaving researchers little choice but to write bespoke scripts (Lee 2014). This situation has improved recently, with the development of the R packages `sExtinct` (Clements 2013) for running various EDEs (e.g., Robson and Whitlock 1964, Strauss and Sadler 1989, Solow 1993, Roberts and Solow 2003, McInerny et al. 2006) and `spatExtinct` (Carlson et al. 2018) for applying the Bayesian approach originally developed by Solow and Beet (2014). An Excel spreadsheet-driven implementation of the Thompson et al. (2013) Bayesian model was also developed by Lee (2014), which can also account for survey effort. The `sExtinct` code only runs 'traditional' EDEs (no uncertainty), whereas the Bayesian methods categorize reliability into types (e.g., unambiguous or ambiguous). Our R script is philosophically different to the Bayesian approach. It is frequentist, and requires specification of reliability scores rather than priors for sighting types, but harmonizes with the `sExtinct` R package because it can make use of its EDE functions (or any other user-defined EDE function).

The decision of how (or if) to account for uncertainty in sighting records remains a topic of contention (Solow et al. 2012). The few 'Lazarus' and 'definitely extinct' species that we assessed using various EDEs (Table 1) suggest that incorporating mixed-certainty sighting records generally improved decision making; however, resolving this question satisfactorily will require an in-depth investigation that is beyond the scope of the current work. What is undeniable is that the lack of consensus on the treatment of uncertainty in sighting records has led to inconsistency in published extinction dates (particularly for controversial species) and subsequent indecision as to which conservation interventions are required (Roberts and Jarić 2016). This has consequential real-world impacts; if a species is proclaimed as extinct and then rediscovered, this erodes researcher credibility, whereas in the converse case, limited funds for conservation might be misallocated (Akçakaya et al. 2017).

An obvious benefit of the BBJ method, or that of Jarić and Roberts (2014), is that there is no need for an arbitrary decision on reliability thresholds. A competing constraint is that assignment of reliability scores is subjective and lacks evidentiary standards, unless sighting data can be validated against some calibration standard (e.g., a test of observer reliability in identifying a bird song to species level, or correctly confirming a sighting, based on a known [experimental] target). However, this is also the case for the threshold approach, where observations are somewhat arbitrarily accepted or rejected based on perceived quality. The reliability ratings used by BirdLife International (see Methods) at least has the advantage of representing a consistent standard that has been developed by an organisation aware of the issues with (and frequency of) misidentification by expert and non-expert bird watchers, and so seems to offer a reasonable *de facto* standard. However, we also recommend that any analysis using sighting reliabilities undertakes a sensitivity analysis, such as lower, upper bounds, to judge the degree to which the conclusion is influenced by this choice. Indeed, it has been argued (McKelvey et al. 2008, Roberts et al. 2010) that flexibility in assigning reliabilities is usually warranted, given knowledge gaps in a species' life-history traits and environmental context, and historical or social biases (Lee et al. 2017, Brook et al. 2018).

Ultimately, the development of a more rigorous, explicit and repeatable framework for dealing with mixed-certainty observational records is critical; the approach we have developed here provides a tool that moves us closer to this goal. But regardless of the method used to account for uncertainty, it is hard to dispute the scientific need. Future research will likely focus on improved ways to incorporate spatio-temporal heterogeneity in sightings (Carlson et al. 2018), sighting probability (Lee et al. 2017), and survey effort (Gotelli et al. 2012, Thompson et al. 2017). Such areas of development should, in principle, be readily incorporated within our general framework.

ACKNOWLEDGEMENTS

This work was funded by Australian Research Council grants FL160100101 and CE170100015 to BWB. We thank Diogo Veríssimo for his help obtaining sighting records. IJ is sponsored by the J.E. Purkyně Fellowship of the Academy of Sciences of the Czech Republic and funding from the Alexander von Humboldt Foundation. The OLE function script was derived from Clements (2013). We thank two anonymous referees for their critical comments and suggestions, and Eli Holmes for her thorough and incisive comments, which substantially improved the final paper.

REFERENCES

Akçakaya, H. R., D. A. Keith, M. Burgman, S. H. M. Butchart, M. Hoffmann, H. M. Regan, I. Harrison, and E. Boakes. 2017. Inferring extinctions III: A cost-benefit framework for listing extinct species. Biological Conservation **214**:336-342.

Boakes, E. H., T. M. Rout, B. Collen, and D. Hodgson. 2015. Inferring species extinction: the use of sighting records. Methods in Ecology and Evolution **6**:678-687.

Brook, B. W., S. R. Sleightholme, C. R. Campbell, and J. C. Buettel. 2018. Deficiencies in estimating the extinction date of the thylacine with mixed certainty data. Conservation Biology **32**:1195-1197.

Carlson, C. J., K. R. Burgio, T. A. Dallas, and A. L. Bond. 2018. Spatial extinction date estimation: a novel method for reconstructing spatiotemporal patterns of extinction and identifying potential zones of rediscovery. bioRxiv:279679; https://doi.org/279610.271101/279679.

Clements, C. 2013. sExtinct R Package. https://CRAN.R-project.org/package=sExtinct.

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

Gotelli, N. J., A. Chao, R. K. Colwell, W. H. Hwang, and G. R. Graves. 2012. Specimen-based modeling, stopping rules, and the extinction of the Ivory-billed Woodpecker. Conservation Biology **26**:47-56.

Jarić, I., and D. L. Roberts. 2014. Accounting for observation reliability when inferring extinction based on sighting records. Biodiversity and Conservation **23**:2801-2815.

Lee, T. E. 2014. A simple numerical tool to infer whether a species is extinct. Methods in Ecology and Evolution **5**:791-796.

Lee, T. E., S. A. Black, A. Fellous, N. Yamaguchi, F. M. Angelici, H. Al Hikmani, J. M. Reed, C. S. Elphick, and D. L. Roberts. 2015. Assessing uncertainty in sighting records: an example of the Barbary lion. PeerJ **3**:e1224.

Lee, T. E., D. O. Fisher, S. P. Blomberg, and B. A. Wintle. 2017. Extinct or still out there? Disentangling influences on extinction and rediscovery helps to clarify the fate of species on the edge. Global Change Biology **23**:621-634.

Lee, T. E., M. A. McCarthy, B. A. Wintle, M. Bode, D. L. Roberts, M. A. Burgman, and J. Matthiopoulos. 2014. Inferring extinctions from sighting records of variable reliability. Journal of Applied Ecology **51**:251-258.

McInerny, G. J., D. L. Roberts, A. J. Davy, and P. J. Cribb. 2006. Significance of sighting rate in inferring extinction and threat. Conservation Biology **20**:562-567.

McKelvey, K. S., K. B. Aubry, and M. K. Schwartz. 2008. Using anecdotal occurrence data for rare or elusive species: the illusion of reality and a call for evidentiary standards. BioScience **58**:549-555.

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

Roberts, D. L., C. S. Elphick, and J. M. Reed. 2010. Identifying anomalous reports of putatively extinct species and why it matters. Conservation Biology **24**:189-196.

Roberts, D. L., and I. Jarić. 2016. Inferring extinction in North American and Hawaiian birds in the presence of sighting uncertainty. PeerJ **4**:e2426.

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

Robson, D. S., and J. H. Whitlock. 1964. Estimation of truncation point. Biometrika **51**:33-39.

Rout, T. M., D. Heinze, and M. A. McCarthy. 2010. Optimal allocation of conservation resources to species that may be extinct. Conservation Biology **24**:1111-1118.

Saltre, F., B. W. Brook, M. Rodriguez-Rey, A. Cooper, C. N. Johnson, C. S. M. Turney, and C. J. A. Bradshaw. 2015. Uncertainties in dating constrain model choice for inferring extinction time from fossil records. Quaternary Science Reviews **112**:128-137.

Solow, A., W. Smith, M. Burgman, T. Rout, B. Wintle, and D. Roberts. 2012. Uncertain sightings and the extinction of the Ivory-billed Woodpecker. Conservation Biology **26**:180-184.

Solow, A. R. 1993. Inferring extinction from sighting data. Ecology **74**:962-964.

Solow, A. R., and A. R. Beet. 2014. On uncertain sightings and inference about extinction. Conservation Biology **28**:1119-1123.

Strauss, D., and P. M. Sadler. 1989. Classical confidence-intervals and Bayesian probability estimates for ends of local taxon ranges. Mathematical Geology **21**:411-421.

Thompson, C. J., V. Koshkina, M. A. Burgman, S. H. M. Butchart, and L. Stone. 2017. Inferring extinctions II: A practical, iterative model based on records and surveys. Biological Conservation **214**:328-335.

Thompson, C. J., T. E. Lee, L. Stone, M. A. McCarthy, and M. A. Burgman. 2013. Inferring extinction risks from sighting records. Journal of Theoretical Biology **338**:16-22.

SUPPORTING INFORMATION

Code (written as scripts for Program R v3.5) to implement the method and all examples reported in the paper is supplied as `DataS1.zip`, and also accessed on the paper's GitHub site: https://github.com/BWBrook/extdyn. Also included are a detailed justification of the basis for the *ad hoc* re-sampling method (Appendix S1: Section S2), further examples of application with case studies (Appendix S1: Section S2) and a sensitivity analysis using simulated data (Appendix S1: Section S3).

TABLE 1. Estimated year of extinction, based on the Solow (1993) equation (S93) and the Roberts and Solow (2003) OLE, for four example species, based on a) physical evidence only (with any ambiguous sightings being rejected), and b) including mixed-certainty sighting records, based on the re-sampling method described in this paper.

		a) Physical evidence only		b) BBJ method	
Type	Species	S93	OLE	S93	OLE
Lazarus	NP	1956 (p_{1989}=0.010)	1941	1998 (p_{1989}=0.565)	2029 (p_{1989}=0.648)
	SB	1970 (p_{1960}=0.063)	1906	1925 (p_{1960}=0.000)	1936 (p_{1960}=0.013)
Extinct	BC	2089 (p_{2014}=0.652)	2009	2035 (p_{2014}=0.999)	2014 (p_{2014}=0.075)
	AG	2014 (p_{2010}=0.059)	1977	1999 (p_{2010}=0.001)	2002 (p_{2010}=0.017)

Notes: The top two rows show 'Lazarus' species that were thought extinct and then subsequently rediscovered (NP = night parrot; SB = noisy scrub bird). The bottom two rows are species where extinction has been 'confirmed' by intensive surveys (BC = Bramble Cay melomys; AG = Alaotra grebe). In parentheses are shown the model-estimated probability of persistence for the year prior to rediscovery (for the Lazarus species) or the date when extinction was confirmed (the year they were considered 'definitely extinct').

TABLE 2. Mean absolute difference between the true (T_{ext}) and the estimated time of extinction (TE) for simulated sighting records in different scenarios (lower values indicate better performance), based on the use of S93 (Solow 1993) and OLE (Roberts and Solow 2003) EDE models and two extinction modes (sudden versus gradual) under: (a) exclusion of sightings which have a reliability below a threshold, which we varied from 0.8 to 0, and (b) BBJ, our new re-sampling approach.

Sighting probability	EDE Method	a. Sighting reliability with threshold exclusion					b. BBJ
		0.8	0.6	0.4	0.2	0	
Constant	S93	8.9	8.2	9.1	10.2	11.3	8.3
	OLE	15.9	13.2	13.3	13.8	14.2	13.1
Declining	S93	9.9	8.4	8.0	8.0	8.2	7.2
	OLE	17.3	12.8	11.7	11.3	11.5	10.4

FIG. 1. Results of applying extinction-date estimators to four well-known and possibly extinct species for which sighting records include observations of mixed certainty. Shown are both a previous analytical approach (JR14, based on Jarić and Roberts 2014) that implements the Solow (1993) model (S93), and a computational re-sampling method (BBJ; this paper), that can provide estimates for any model—shown here is last appearance date (LAD), S93, and the ordinary least squares method of Roberts and Solow (2003), OLE. The colours represent the use of different probabilities for the record types (in this example being physical, expert or controversial), with the probabilities being, respectively: P-only = 1,0,0; Extreme = 0.99,0.5,0.01 and BL-M (based on the mid-point probabilities selected by BirdLife International, see Lee et al. 2014) = 0.85, 0.7, 0.25. The dots represent point (median) estimates and dashed lines are the upper 95% confidence bounds (the lower bounds are not of interest in this context). The proportion of observations of each species falling into the physical, expert and controversial types respectively were: *Panthera leo leo* = 0.12, 0.47, 0.41; *Lipotes vexillifer* = 0.47, 0.49, 0.04; *Campephilus principalis* = 0.32, 0.25, 0.43; *Hemignathus lucidus* = 0.3, 0.03, 0.67.