Bayesian updating to estimate extinction from sequential observation data

Colin J. Thompsona, Saritha Kodikarab, Mark A. Burgmanc, Haydar Demirhanb and Lewi Stoneb,d

*aDepartment of Mathematics and Statistics, University of Melbourne, 3010 Australia*

*bSchool of Sciences, Mathematical Sciences Discipline, RMIT University, Melbourne, Australia*

*cCentre for Environmental Policy, Imperial College London, UK, and School of BioSciences, University of Melbourne, 3010 Australia*

*dBiomathematics Unit, Department of Zoology, Faculty of Life Sciences, Tel-Aviv University, Israel*

**Corresponding author**: Prof. Colin Thompson, Department of Mathematics and Statistics, University of Melbourne, Victoria, 3010 Australia

Phone: +61 3 8344 5550

Email: [markab@unimelb.edu.au](mailto:markab@unimelb.edu.au)

Abstract

Several new approaches to estimating the probability that a species is extinct have emerged recently. Different foundational assumptions can lead to different interpretations of data and potentially to different conclusions. To explore the implications of alternative formulations, here we develop and illustrate a Bayesian Updating method for inferring extinction based on records of observations and surveys. We illustrate how it combines incidental sightings and surveys with a data set for the Alaotra Grebe, showing how estimates of extinction may be updated as new data arise, providing a means for managers to reassess priorities for survey and management dynamically.

*Key-words:* extinction, threatened species. observations, surveys, Bayes rule, dynamic updating, Bayes factors

1. **Introduction**

Extinctions in ecology are important for many reasons, not least because they represent the ultimate expression of many human impacts on the natural world. However, it is often difficult to know when the last member of a species has died. Instead, extinction is inferred from absences of incidental sightings and from dedicated surveys which fail to detect the target species. The extinction or otherwise of species can be controversial (see Brook et al. 2018, Carlson et al. 2018 for discussion of the status of the thylacine, and Collins 2017, for a discussion of the kinds of data available to support inferences about the status of the Ivory Billed Woodpecker). Incorrect inferences may lead to actions to protect species that are already extinct, or to a failure to act when it may have been effective to do so (Akcakaya et al. 2017).

The problem of inferring possible extinction of a species from sighting records (data) has been much studied and debated (Lee, Bowman, & Roberts, 2017; Solow & Beet, 2014; Thompson et al. 2013), beginning with the seminal work of Solow (1993 a, b, 2005). Many models have been proposed that use different assumptions and data, and which may give incompatible and even contradictory (i.e. inconsistent) estimates for probabilities of extinction (Thompson et al. 2017). For example, the model developed by Solow and Beet (2014; see Bond et al. 2018 for an application) does not use information on detection probability and survey effort. Thompson et al. (2017) proposed a non-Bayesian linear model (LM) for the extinction problem to account for such data when the data are sequential in time. In this approach, the probability of extinction in year , is updated to determine as new sighting data come to hand year by year. In their LM model there are no Bayes Factors, rules of thumb or priors.

In the present article we propose an alternative Bayesian approach for circumstances in which sighting data are sequential in time. Known as Bayesian Updating (O’Hagan, 2017), probabilities of extinction are updated year-by-year as new data come to hand, by taking the (Bayesian) “posterior” in year to be the “prior” in year . This results in a non-linear iterative model (Bayesian Updating model; BU) with a cumulative Bayes Factor which is also updated year-by-year.

In the following section we present a detailed description of the BU model including an exact expression for the probability that the species is extant in year in terms of an “initial” and cumulative Bayes Factor . The choice of is discussed and explicit expressions are given for yearly Bayes Factors in terms of parameters for recordings and unsuccessful surveys (described in detail below). In section 3 we present a methodology for dealing self-consistently with rules of thumb and probability thresholds for the BU model.

As a case study in section 4, we consider Alotra Grebe (*Tachybaptus rufolavatus*) which was the subject of a detailed analysis using our LM in Thompson et al. (2017). Our results are summarized and discussed in the final section.

1. **Bayesian Updating**

Consider a period of *T* consecutive years of sequential data . To be specific we consider a species which is either extant (*x*) or extinct (*e*) at the beginning of any year in the record period (1*, T*) with probabilities and. The sequential data {} in this case represent sighting states (in year *t*) which could take many forms including “recordings (*r*)” (photographs, sounds, specimens, …), or unsuccessful surveys (*u or u’*) covering some fraction of the species’ habitat. Thus, ‘*r’* represents a successful sighting / survey. If there are no sightings during a dedicated survey, then it is classified as an active unsuccessful survey (*u*). In the absence of a dedicated survey, it is assumed that there may still be unplanned surveys by interested professionals or amateur ecologists. These kinds of unsuccessful surveys are referred to as passive unsuccessful surveys (*u’*).

In any given year, Bayes Rule states that with sighting state ‘*s*’ in that year, the posterior (conditional probability) is given by

There is of course a similar equation for, in terms of the inverse conditional probabilities and and the priors and . The ratio of these two equations gives

(2)

where

(3)

is the Bayes Factor for that year.

Note that we only have one sighting record in the period (1, *T*) specified by the given sequence { of sighting states. Uncertainty in this case is embodied in the unknown and uncertain values of the priors and the Bayes Factors eqn (3), the latter depending on the independently of the prior(s).

Proceeding year by year we propose a simple updating methodology to determine the probability that the species is extant in year from and the Bayes Factor in year. Specifically, in eqn (2) we take the priors in year to be

and (4)

with Bayes factor in year given by, from eqn (3),

. (5)

We then update to year by taking the posteriors in year from eqn (2) to be the priors for year, i.e. from eqns (2), (4) and (5)

(6)

We refer to the iterative rule eqn (6) as Bayesian updating. If we now substitute the second equation of eqn (4) into eqn (6) we obtain a linear iterative equation for which can be readily solved and inverted to give

(7)

where

(8)

may be interpreted as a cumulative Bayes Factor.

To implement the Bayes updating model (BU) eqn (7), we need to assign values to the initial and the yearly Bayes Factors in eqn (5).

**Choosing an initial**

In the conventional Bayes Rule of eqn (1) a prior of unity (, i.e. ) implies a posterior of unity (). Similarly, in the BU analysis, an initial from eqn (7) implies for all . However, *P*(*X*1) is uncertain and we specify a range of values for, assuming a uniform (or some other) distribution over this range. For example, we may replace by its mid-point value over a specified range. If one assumes a value of between 0.25 and 0.75 one could choose the average value of 0.5, giving equal weight to the species being extant or extinct. We will adopt such mid-point choices of BU model parameters, described below. This approach is discussed in greater detail in Thompson et al. (2013).

**Calculating the yearly Bayes Factor**

The following formulas use parameters that may be estimated from data or by ornithological specialists familiar with the species in question (see Thompson et al. 2017). There are two main sighting states are recordings (*r*) and unsuccessful surveys (*u* active or *u’* passive).

1. For a recording year (*r*) there is one parameter:

the probability that the recorded species is correctly identified.

Then by definition in an *r* year is the probability that the species is incorrectly identified. In addition it is clear that. Hence from eqn (3)

. (9)

1. For an active unsuccessful survey year (*u*) there are two parameters:

= the probability that the species could have been reliably identified and recorded. Note that following Thompson (2017), where is the probability that the species could have been reliably identified in the survey if it had been recorded. is the probability that the species would have been recorded in the survey.

= the proportion of the species habitat within its likely range that was surveyed.

Then by definition in a *u* year is the probability that the survey was unsuccessful. In addition, it is also clear that. Hence from eqn (3)

. (10)

In passive unsuccessful survey years (*u*’), and in eqn (10) are replaced by their primed counterparts.

To summarize, the Bayes Factor in year can be expressed, from eqns (5), (9) and (10) as,

= (11)

1. **Rules of Thumb and Probability Thresholds**

In our basic eqn (7), is sensitive to the choice of the initial just as posteriors in conventional Bayesian analysis are sensitive to choices of priors. In the latter case, it is common practice to deal directly with Bayes Factors ( which are independent of priors, and to specify rules of thumb values (*R*), arguing that a *B* exceeding *R* provides “substantial evidence for extinction”.

An equivalent situation applies to the BU model where the cumulative Bayes Factor given by eqn (8) is independent of the initial. The problem then is to specify values for . Depending on applications values ranging from 3 to 100 (or more) have been suggested (Jeffreys, 1961; Solow & Beet, 2014). Ideally of course one would like to estimate probabilities of extinction, e.g. for BU from eqns (7) and (8), and to use probability threshold lower bounds on (e.g. 0.95) to make decisions regarding extinction of species. In any event, one is left with the problem of choosing. As an example for BU if one chooses in eqn (7) one may need a value of of at least 20 or so (rising to 60 if one chooses) to achieve a probability threshold value for of 0.95. That is, with, one needs in order for

A novel feature of the BU model is that the Bayes Factor eqn (8) is cumulative and from eqn (11) decreases in “*r*” years and increases in “*u, u’*” years. In particular, if we set to the last “*r*” year in a sighting record, is an increasing function of. One way to proceed is then to calculate for successive until one reaches where first exceeds a specified rule of thumb (e.g. 20) at which point, the probability of extinction exceeds a given probability threshold (e.g. 0.95).

In the following section we consider Alaotra Grebe as a case study for our BU model, taking us back to our previous analysis of Grebe data using an alternative linear model in Thompson et al. (2017).

1. **Case Study**

The water bird, the Alaotra Grebe (*Tachybaptus rufolavatus*) was once endemic to Lake Alaotra and surrounding lakes in Madagascar. It was last observed in 1988 and as a result of several large scale unsuccessful surveys (*u’*) over the period 1989-1997, it was inferred to be extinct by the mid 1990’s by several research groups (Thompson et al. 2017, Keith et al. 2017).

In the following analysis we use observation (records) data for the Alaotra Grebe collected by BirdLife International as given in Thompson et al. (2017). Table 1 presents and adapts some of those data. We note firstly that over the 18 year period from 1970-1988 there were 7 “*r*” years and 11 “*u* ” years. Using midpoint values in *r* years and midpoint values for and in *u’* years (from Table 1 in Thompson et al. (2017), i.e. ) and an initial of 0.5 in 1970, we find is slightly less than 1 in 1988 then decreases markedly in subsequent *u* years in accord with the results show in Figs 1, 2, 3 of Thompson et al. (2017). We therefore focus our attention on years following 1988 where there were only unsuccessful (*u, u’*) surveys. In the following analysis we thus take to be 1988.

In view of the above remarks, we initiate the BU model in 1990 and take in that year. We then need to check for consistency of these assumptions as described in the previous section. Firstly, however, from Table 1 in Thompson et al. (2017) we have *u* years in 1989, 1990, 1993, 1994, 1997, 1998, 1999 with midpoint parameters values of and giving a Bayes Factor in those years from eqn (11) of . In *u*’ years after 1988, midpoint parameter values of and give a Bayes Factor from eqn (11) of. Cumulative Bayes Factors from eqn (8) then take respective values from 1990 to 1994 of 2.837, 2.850, 2.864, 8.124, 23.05, suggesting from a rule of thumb value of that extinction occurred on or before 1995 (see Table 1).

Table 1. Passive and active survey input data for the Alaotra Grebe data commencing in 1990. The low and high values for the parameter are taken from Thompson et al. (2017). As discussed in the text, here we use the midpoint values for each parameter, and the fact that (as discussed in main text).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Start Year | P(Xt) |  |  |  |  |  |  |  |  |  | |  | |  |
| 1990 | 0.5 |  |  |  |  |  |  |  |  |  | |  | |  |
|  |  |  |  |  |  |  |  |  |  |  | |  | |  |
| **Passive Surveys (u')** | | | | | | | | | | | | | | |  |
| ɛ' | | | p'(i) | | | p'(r) | | | ɛ'p'(ri) | | | | | |  |
| Low | Mid | High | Low | Mid | High | Low | Mid | High | Low | | Mid | | High | |  |
| 0.00 | 0.03 | 0.05 | 0.10 | 0.38 | 0.65 | 0.40 | 0.50 | 0.60 | 0.0000 | | 0.0047 | | 0.0195 | |  |
|  |  |  |  |  |  |  |  |  |  |  | |  | |  |
|  |  |  |  |  |  |  |  |  |  |  | |  | |  |
| Year | **Active Surveys (u)** | | | | | | | | | | | | | |
| ɛ | | | p(i) | | | p(r) | | | ɛp(ri) | | | | |
| Low | Mid | High | Low | Mid | High | Low | Mid | High | Low | | Mid | | High |
| 1990 | 0.80 | 0.88 | 0.95 | 0.90 | 0.93 | 0.95 | 0.70 | 0.80 | 0.90 | 0.50 | | 0.65 | | 0.81 |
| 1993 | 0.80 | 0.88 | 0.95 | 0.90 | 0.93 | 0.95 | 0.70 | 0.80 | 0.90 | 0.50 | | 0.65 | | 0.81 |
| 1994 | 0.80 | 0.88 | 0.95 | 0.90 | 0.93 | 0.95 | 0.70 | 0.80 | 0.90 | 0.50 | | 0.65 | | 0.81 |
| 1997 | 0.80 | 0.88 | 0.95 | 0.90 | 0.93 | 0.95 | 0.70 | 0.80 | 0.90 | 0.50 | | 0.65 | | 0.81 |
| 1998 | 0.80 | 0.88 | 0.95 | 0.90 | 0.93 | 0.95 | 0.70 | 0.80 | 0.90 | 0.50 | | 0.65 | | 0.81 |
| 1999 | 0.80 | 0.88 | 0.95 | 0.90 | 0.93 | 0.95 | 0.70 | 0.80 | 0.90 | 0.50 | | 0.65 | | 0.81 |
| 2000 | 0.70 | 0.80 | 0.90 | 0.90 | 0.93 | 0.95 | 0.70 | 0.80 | 0.90 | 0.44 | | 0.59 | | 0.77 |
| 2004 | 0.80 | 0.88 | 0.95 | 0.90 | 0.93 | 0.95 | 0.70 | 0.80 | 0.90 | 0.50 | | 0.65 | | 0.81 |
| 2009 | 0.80 | 0.88 | 0.95 | 0.90 | 0.93 | 0.95 | 0.70 | 0.80 | 0.90 | 0.50 | | 0.65 | | 0.81 |

The data in Table 1 were used to construct Figures 1 a and 1b. They show the cumulative Bayes Factor for the likelihood that the species is extinct, together with the probability that the species is extant, which declines in the absence of successful surveys and sightings from the initial value of 0.5 in 1990 to a value close to zero by 1998.

Figures 1 a, b. Cumulative Bayes Factors and probabilities of extinction generated sequentially from the data in Table 1.

Values for the model from (1989) to (1995) derived from eqn (7) for BU (assuming an initial of 0.5 in 1990 and the above Bayes Factors). The values for show probabilities for extinction by 1995 of , exceeding a notional probability threshold value of 0.95. There are many possible variations on the model outputs above. For example, if we initiate yearly Bayes Factors, from 1989 we obtain a cumulative Bayes Factor in 1994 of 65.7 (suggesting a rule of thumb value of, say, 65).

1. **Discussion**

Inferring extinction probabilities is important in ecology because these inferences affect reporting on the state of the environment, and decisions about priorities for surveys, actions to abate threats, and to establish and manage protected areas (see Thompson et al. 2017). The IUCN (2012) defines a taxon as extinct “when there is no reasonable doubt that the last individual has died. A taxon is presumed extinct when exhaustive surveys in known and/or expected habitat, at appropriate times (diurnal, seasonal, annual), throughout its historic range have failed to record an individual. Surveys should be over a time frame appropriate to the taxon’s life cycle and life form”. We have described an approach to inferring possible extinction of a species from sighting records that allows for opportunistic sightings and dedicated surveys, accounting for the period of observation and the extent of habitat surveyed.

Our approach provides a means for quantifying what is meant by ‘no reasonable doubt’ and a basis for consistent allocation of resources to mitigate extinction risks (Akcakaya et al. 2017). Its particular utility in this regard is that it provides a means of reassessing estimates and priorities as new information arises. The model is very simple to implement in a standard spreadsheet, facilitating its adoption for routine application in organisations that use observation data to support evidence-based decisions regarding actions to protect species and their habitats. It should also provide a measure of comfort for those who have used sighting models in controversial circumstances (Carlson et al. 2018), providing an opportunity for a more comprehensive, integrated interpretation of data.

Different foundational assumptions can lead to different interpretations of data and in some instances, different conclusions. It is important that such differences be reconciled. Here, we have developed and illustrated a Bayesian Updating method for evaluating the probability that a species is extinct, based on a record of observations and surveys. We have shown that it is consistent with a non-Bayesian model and illustrated its use with a data set to arrive at conclusions which accord with the non-Bayesian model applied to the same data.

As with the model developed by Thompson et al. (2017), the Bayesian Updating model may be used to explore hypothetical scenarios, or to test ideas about investments in surveys. For instance, this model will be useful if questions arise regarding the trade-offs between more extensive surveys, or new technologies that are more likely to detect a species when it is present. Manipulations of the model’s parameters will reveal whether potential investments will contribute substantially to the estimated probability that a taxon is extinct. Likewise, alternative targeted survey strategies or training scenarios may be assessed (Akcakaya et al. 2017, Thompson et al. 2017). Perhaps most importantly, the methods outlined here indicate how acceptable levels of uncertainty may be quantified so that decisions about the allocation of resources may be made transparently and consistently.

**Acknowledgements**

We thank the referees and editors for excellent comments that improved the quality of the manuscript.

## References

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

Bond, A.L., Carlson, C.J. & Burgio, K.R. (2018) Local extinctions of insular avifauna on the most remote inhabited island in the world. *Journal of Ornithology*. https://doi.org/10.1007/s10336-018-1590-8

[Brook](https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorStored=Brook%2C+Barry+W), B.W., [Sleightholme](https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorStored=Sleightholme%2C+Stephen+R), S.R., [Campbell](https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorStored=Campbell%2C+Cameron+R), C.R. & Buettel, J.C.. 2018. Deficiencies in estimating the extinction date of the thylacine with mixed certainty data. *Conservation Biology* 32, 1195-1197.

Carlson, C.J., Bond, A.L. & Burgio, K.R. (2018). Re-evaluating sighting models and moving beyond them to test and contextualize the extinction of the thylacine. *Conservation Biology* 32, 1198-1199.

Collins M.D. (2017) Video evidence and other information relevant to the conservation of the Ivory-billed Woodpecker (*Campephilus principalis*). *Heliyon* 2017 Jan 1;3(1):e00230.

IUCN. (2012). IUCN Red List Categories and Criteria: Version 3.1. Second edition. Gland, Switzerland and Cambridge, UK: IUCN

Jeffreys, H. (1961) Theory of probability. *Oxford University Press, Oxford*.

Keith, D. A., Butchart, S. H., Regan, H. M., Harrison, I., Akçakaya, H. R., Solow, A. R. & Burgman, M. A. (2017). Inferring extinctions I: A structured method using information on threats. *Biological Conservation*, *214*, 320-327.

Lee, T. E., Bowman, C. & Roberts, D. L. (2017). Are extinction opinions extinct ? *PeerJ*, *2017*(8). https://doi.org/10.7717/peerj.3663

O’Hagan, A. (2017). Fractional Bayes Factors for Model Comparison. *Journal of the Royal Statistical Society*, *57*(1), 99–138. Retrieved from http://www.jstor.org/stable/2346088

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

Solow, A.R. (1993b). A simple test for change in community structure. *Journal of Animal Ecology,* **62**, 191-193.

Solow, A. R. & Beet, A. R. (2014). On uncertain sightings and inference about extinction. *Conservation Biology*, *28*(4), 1119–1123. https://doi.org/10.1111/cobi.12309

Thompson, C. J., Lee, T. E., Stone, L., McCarthy, M. A. & Burgman, M. A. (2013). Inferring extinction risks from sighting records. *Journal of Theoretical Biology*, *338*(August), 16–22. https://doi.org/10.1016/j.jtbi.2013.08.023

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

## 