Inferring Extinction Year using a Bayesian Approach

Saritha Kodikara[1], Haydar Demirhan[1], Andrew Solow[2], Yan Wang[1] and Lewi Stone[1,3]

[1]School of Science, Mathematical Sciences, RMIT University, Melbourne, Australia. Email:sarithakalhari.kodikara@rmit.edu.au

[2]Woods Hole Oceanographic Institution, Woods Hole, MA 02543, U.S.A

[3]Biomathematics Unit, School of Zoology, Faculty of Life Science, Tel-Aviv University, Israel

30th January 2021

This is the pre-peer reviewed version of the following article: (Kodikara *et al.* 2020), which has been published in final form at https://doi.org/10.1111/2041-210X.13408. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

Abstract

1. Species sighting records are combined with statistical models to infer whether an endangered species might have become extinct or whether instead it has just gone unobserved for a lengthy period of time. The challenging part in developing these models lies in the inclusion of uncertain sightings.

2. We propose a Bayesian hierarchical approach to infer the 1 extinction time of a

species based on historical sighting records which may be either certain or uncertain. The posterior distribution for extinction time is evaluated using the likelihood of sighting data and non informative priors for model parameters. All the models discussed in this paper are implemented in JAGS, a program for analyzing Bayesian models using Markov Chain Monte Carlo (MCMC) simulation.

3. A general methodology is presented, and it is then applied to the sighting record of the Ivory-billed Woodpecker (IBW) *(Campephilus principalis)* . It was found that the IBW most likely went extinct between 1940 and 1945, a little after the date of the last certain sighting. Interestingly, for the IBW dataset, the inclusion of uncertain sightings did not significantly change the inference about the extinction date. We explore the important role of the last certain sighting when estimating the extinction date.

4. When estimating the extinction date of a species it is important to understand the role of the last certain sighting. If there are no uncertain sightings in the sighting record then the species is highly likely to go extinct soon after the last certain sighting. But when there are uncertain sightings a species is likely to go extinct closer to the **Let's discuss this** last certain sighting or to the time point where the uncertain sightings drops to a lower rate. Applying our analysis to a real dataset, we find that the IBW most likely went extinct in 1940, despite recent controversial claims that it was sighted in 2004.

Key words: Bayesian modeling, Extinction probability, Highest posterior density interval, Markov Chain Monte Carlo, Sighting record, Uncertain sightings

1 Introduction

Clear signs are emerging that any further loss of critically endangered species might tip the world towards another mass extinction event (Barnosky *et al.* 2011). These extreme events have likely only occurred five times in the past 540 million years (Barnosky *et al.* 2011; Pimm *et al.* 2014). As such, there is concern that the diversity and complexity of life on Earth may well again be on a dangerous downward spiral. It also highlights the need to correctly monitor and model the current extinction status of species on planet Earth and carefully assess the fragility of potentially endangered species. An incorrect classification of a species as extinct can lead to failure in conserving a threatened species (Lee *et al.* 2014; Thompson *et al.* 2013). On the other hand, it is also undesirable to classify a species as extant when it is actually extinct, as it can lead to misallocation of research energy and funds (Thompson *et al.* 2013; Lee *et al.* 2014; Akçakaya *et al.* 2017; Keith *et al.* 2017).

In practice, it is extremely difficult to determine whether a species has gone extinct or has just remained unobserved (Akçakaya *et al.* 2017; Keith *et al.* 2017; Thompson *et al.* 2017). But it only requires one certain sighting to prove that a species is extant. A recent example of an erroneously inferred extinction is the Aldabra banded snail *Rhachistia aldabrae*. Gerlach (2007) announced that these snails went extinct as a result of short-term climate change, as no recent shell or live specimen was sighted after 1997. This was the case even after systematic and exhaustive surveys specifically aimed at finding the snail in 2005 and 2006. Nevertheless, the snail surprisingly reappeared in 2014, when the rediscovery was publicised by the Seychelles Island Foundation (Battarbee 2014).

Historical sighting records are often the only available data for rare or poorly studied species, and thus the main information available to work with for quantitative assessment of extinction. Palaeobiologists first introduced the general idea of using sighting records to estimate the time of extinction (Strauss & Sadler 1989; Marshall 1990), while Solow (1993) applied it for the first time within the field of conservation biology. Solow (Solow 1993) developed a Bayesian approach to derive an equation for expressing the survival probability of a species based on sightings over a series of time units.

Rivadeneira *et al.* (2009) pointed out that most of the statistical methods for assessing species extinction before 2009 assumed that all sightings were valid with complete certainty.

However, the conclusions about extinction are sensitive to the inclusion or exclusion of sightings that may be "uncertain." This spurred modellers to examine better what happens when uncertainty might be attached to the validity of sightings. Roberts *et al.* (2010) noted that inferences from models including uncertain sightings differ significantly from those obtained by models omitting this information. Several studies were developed to incorporate probabilities of reliability or sighting validity for each sighting into the model development (Jarić & Roberts 2014; Lee *et al.* 2014; 2015; 2017), as well as expert opinion (Lee *et al.* 2015).

When analyzing the different approaches in (Solow & Beet 2014), it was found that the final inferences made were particularly sensitive to the different ways of modeling uncertain sightings (Kodikara *et al.* 2018) as suspected by (Solow & Beet 2014). This indicates the need to gain a deeper understanding and familiarity with models that include uncertain sightings. There has been recent interest in developing frameworks that incorporate uncertain sightings (Solow *et al.* 2012; Solow & Beet 2014; Thompson *et al.* 2017).

Our approach is based on working with posterior probabilities rather than the Bayes factor in order to make inferences. There are a few Bayesian studies that calculate the posterior probability of extinction (Alroy 2014; 2016; Solow 2016a) and none of these have dealt with uncertain sightings so far. The models developed here are implemented using JAGS software, which uses a computational Bayesian approach based on MCMC simulations. This approach has become very popular of late, in particular with the availability of statistical packages, such as JAGS, WinBUGS etc.

The remainder of this paper is organized in the following way. Section 2 presents the development of the models. Section 3 explores the models using the sighting records of the IBW. Section 4 examines the sensitivity of the results to uncertain sightings. The paper concludes with the discussion in Section 5.

2 Model Development

Consider a historical sighting record S of a species in which n sightings occurred in years $S = (s_1, ..., s_n)$, as recorded over the full observation period $t = 1, \ldots, T$ years. If the species went extinct during the observation period, then we designate τ_E as the date of the first year following extinction. In this paper, a hierarchical Bayesian approach is developed to infer the extinction time τ_E for a species based on its sighting record S. From this it is possible to infer the probability a species went extinct during the observation period, that is $p(\tau_E \leq T|S)$.

Sightings in S can either be certain or uncertain and this is something that has to be fully taken into account. Note that all certain sightings are taken to be valid given the species has been correctly identified on each sighting date. However, uncertain sightings can either be valid or invalid (since now the species is sometimes incorrectly identified).

Bayesian inference is used to find the posterior probability distribution for parameters of interest (eg., τ_E), based on prior knowledge of the parameter combined with a statistical model of the observed data (likelihood function). This requires working with the well known Bayesian formula:

$$posterior \propto likelihood \times prior \tag{1}$$

Here the prior is our initial knowledge about the parameter of interest. While the posterior is a revised updated version of the prior for which the observed data has been into account via the likelihood.

Two distinct modeling approaches are developed. The first is appropriate for sighting records that consist of certain sightings only. The second includes uncertain sightings into the model.

We next discuss the development of the likelihood in each of these models. From there we show how the likelihood and prior specification is used to obtain the posterior distribution according to Equation 1.

2.1 Model 1 - Certain sightings only

Formulation of the likelihood

First consider a historical sighting record S of a species in which all n sightings recorded are certain and occur in years $S = C = (c_1, ..., c_n)$. Thus c_n is the time of the last certain sighting. The model assumes that there is a probability p_c that an extant species can be sighted in any given year. Our goal is to infer the distribution of extinction times τ_E, based on the sighting record data S. Clearly, τ_E must be greater than the last certain sighting c_n. When there are n_c certain sightings, the likelihood for the sighting record S given τ_E and p_c is easily seen to be:

$$p(S|\tau_E, p_c) = p_c^{n_c}(1 - p_c)^{(\tau_E - 1 - n_c)}. \quad (2)$$

Since the full sighting record occurred in the period $(0, T)$, the upper bound for τ_E (the year following extinction) should be $T + 1$. Hence, the likelihood of S given $\tau_E > T$ is found by evaluating $p(S|\tau_E = T + 1, p_c)$. Considering all situations, Equation 2 for the likelihood should be generalized as follows:

$$p(S|\tau_E, p_c) = \begin{cases} 0, & \tau_E \le c_n \\ p_c^{n_c}(1-p_c)^{(\tau_E-1-n_c)}, & c_n < \tau_E \le T \\ p_c^{n_c}(1-p_c)^{(T-n_c)}, & \tau_E > T. \end{cases} \tag{3}$$

The basic set up of Model 1 is identical to the one proposed by Alroy (2014) and the associated paper of Solow (2016b). In the latter study, instead of treating the yearly sighting probability p_c as a parameter of interest, an approach was developed to completely eliminate p_c by treating it as a nuisance parameter, resulting in an analytical solution for the posterior extinction probability. However, for the reasons that will become evident shortly, it is instructive and useful to include p_c.

Prior distributions of model parameters

Assuming that an extant species can become extinct (E) at the beginning of each year with probability θ, the number of years until the species becomes extinct τ_E is characterized by a geometric distribution with parameter θ.

$$p(\tau_E|\theta) = (1-\theta)^{\tau_E-1}\theta, \quad \tau_E = 1, 2, ... \tag{4}$$

The parameter p_c and hyper-parameter θ are assumed to have a standard uniform distribution with the following probability density function,

$$p(\theta) = 1, \quad 0 < \theta < 1. \tag{5}$$

and

$$p(p_c) = 1, \quad 0 < \theta < 1. \tag{6}$$

Based on this framework, we now evaluate the posterior distribution of τ_E.

Posterior distribution

Applying Bayes' rule defined in Equation 1, the posterior distribution for the parameters of interest (τ_E, p_c and θ) is written as the product between the likelihood, priors and hyper-prior as follows:

$$p(\tau_E, p_c, \theta|S) \propto p(S|\tau_E, p_c)p(\tau_E|\theta)p(\theta)p(p_c) \tag{7}$$

where

$p(\tau_E, p_c, \theta|S)$ = posterior distribution for τ_E, p_c and θ given the observed data S;

$p(S|\tau_E, p_c)$ = likelihood function for S given τ_E and p_c;

$p(\tau_E|\theta)$ = prior distribution of τ_E given the hyper-parameter θ and

$p(\theta)$ = hyper-prior distribution of θ.

$p(p_c)$ = prior distribution of p_c.

According to Equation 7, we obtain the MCMC samples for τ_E using JAGS. The model developed in JAGS is specified according to the likelihood function defined in Equation 3 along with the prior and hyper-prior distributions in Equations 4, 6 and 5. Since τ_E is a discrete random variable, the posterior distribution of τ_E describes the probability of occurrence of each value of τ_E. By summing all probabilities that are less than or equal to T in the posterior distribution of τ_E, we can obtain the posterior probability of $p(\tau_E \leq T|S)$ which can be expressed by the following formula:

$$\begin{aligned} p(\tau_E \leq T|S,\theta,p_c) &= \frac{p(\tau_E \leq T|S,\theta,p_c)}{p(\tau_E|S,\theta,p_c)} \\ &= \frac{\sum_{\tau_E=s_n+1}^{T}(1-p_c)^{(\tau_E-1)}(1-\theta)^{\tau_E-1}\theta}{\sum_{\tau_E=s_n+1}^{T}(1-p_c)^{(\tau_E-1)}(1-\theta)^{\tau_E-1}\theta+(1-p_c)^{T}\sum_{\tau_E=T+1}^{\infty}(1-\theta)^{\tau_E-1}\theta}. \end{aligned} \tag{8}$$

A similar formulation was used in (Fader *et al.* 2010; Thompson *et al.* 2013; Alroy 2014). The method discussed in Thompson *et al.* (2013) uses a simple estimate for the probability of sighting a species when it is extant, i.e., dividing the number of years in which there are sightings by the time of the last sighting ($\hat{p}_c = \frac{n}{t_n}$). Our approach, goes beyond this simple method and estimates the parameter p_c using a Bayesian approach.

2.2 Model 2 - Certain and uncertain sightings

Many historical data sets of rare or extinct species contain sightings that are to some degree uncertain. While physical evidence of a species is usually taken to indicate that the species was certainly present during a survey, other evidence is often less certain. Suppose that the certain sightings occur in years $C = (c_1, ..., c_n)$ and uncertain sightings occur in years

$U = (u_1, ..., u_n)$, where c_n and u_n represents the time of the last certain and last uncertain sighting respectively. Then the sighting record S is a combination of both C and U records. Our work assumes that uncertain sightings can only be recorded in years in which there are no certain sightings. In other words there is some "censorship" process that masks the recording of uncertain sightings.

A likelihood for the sighting record S can be constructed similar to the "certain sighting only" model that takes into account the censorship process. Consider first the case $c_n < \tau_E \leq T$, Then, in any year before extinction $t < \tau_E$, a sighting is considered an outcome of a generalized Bernoulli trial where either a certain sighting or an uncertain sighting or no sighting is recorded. For any year after extinction ($t \geq \tau_E$), all uncertain sightings are invalid. Thus a sighting is considered a Bernoulli trial with either an invalid uncertain sighting with probability p_{ui}, or no sighting with probability $1 - p_{ui}$, as outcomes.

Next we discuss how to allow for the censoring process (i.e., no single year can have both certain and uncertain sightings). Recall that for an extant species. the probability of recording a certain sighting in any year during is p_c. Also, it is natural to assume that valid uncertain sightings and invalid uncertain sightings occur independently according to some probabilities, say p_{uv} and p_{ui}. Thus the probability of having an uncertain sighting before τ_E is $p_u = p_{uv}(1 - p_{ui}) + p_{ui}(1 - p_{uv}) + p_{uv} * p_{ui}$. Recall that uncertain sightings are only recorded if there are no certain sightings. Because of this "censoring" process, even though the probability of an uncertain sighting is p_u, the probability of recording it is $(1 - p_c)p_u$. The probability of not recording an uncertain sighting is $(1 - p_c)(1 - p_u)$. (If one prefers to consider the model without the censoring process then the above defined probabilities should be modified accordingly.)

The certain sighting record C consists of n_c sightings, all of which occur prior to τ_E as there cannot be any certain sighting after extinction. Let N_u be the total number of uncertain sightings. The uncertain sighting record t_u consists of $n_u(\tau_E)$ sightings prior to τ_E of uncertain validity, followed by $N_u - n_u(\tau_E)$ sightings after τ_E all of which must be invalid. When $\tau_E > T$, then $n_u(\tau_E) = N_u$. Considering all situations described above, the likelihood $p(S|\tau_E, p_c, p_{ui}, p_{uv})$ can be summarized as:

$$p(S|\tau_E, ...) = \begin{cases} 0, & \tau_E \leq c_n \\ \\ p_c^{n_c} * ((1-p_c)p_u)^{n_u(\tau_E)} & \\ \quad *((1-p_c)(1-p_u))^{\tau_E - 1 - n_c - n_u(\tau_E)} & c_n < \tau_E \leq T \\ \quad *p_{ui}^{N_u - n_u(\tau_E)} * ((1-p_{ui}))^{T-(\tau_E-1)-(N_u - n_u(\tau_E))}, & \\ \\ p_c^{n_c} * ((1-p_c)p_u)^{N_u} & \\ \quad *((1-p_c)(1-p_u))^{T-n_c-N_u} & \tau_E > T. \end{cases} \tag{9}$$

The key notations used in Equation 9 are summarised in Table 1. In Equation 9, we have used the result that the likelihood of counts n_1, n_2 and n_3 arises from a generalized Bernoulli trial with probabilities p_1, p_2 and $p3$ (i.e. $p_1 + p_2 + p_3 = 1$) is $p_1^{n_1} * p_2^{n_2} * p_3^{n_3}$.

Notation	Description
τ_E	Time or date of first year following extinction.
c_n	The date of the last certain sighting.
n_c	The total number of certain sightings.
N_u	The total number of uncertain sightings.
$n_u(\tau_E)$	The number of uncertain sightings prior to τ_E.
p_c	The probability of having a certain sighting in each year.
p_{uv}	The probability of having a valid uncertain sighting in each year.
p_{ui}	The probability of having an invalid uncertain sighting in each year.

Table 1: Notation used in model development

Note that once Model 2 is operational, it is simple to run Model 1 by setting the uncertain sighting probabilities to zero (i.e. $p_{ui} = 0$ and $p_{puv} = 0$).

As discussed in the previous Subsection, τ_E was modelled as a geometric distribution with parameter θ, where the prior for θ was taken to be a uniform(0,1) distribution. All the other parameters (i.e p_c, p_{ui} and p_{uv}) were also assigned a standard uniform prior. Using these prior specifications along with the likelihood in Equation 9, we obtained posterior distributions for all model parameters including, most importantly, τ_E. In any MCMC implementation, we generated 4 chains each with 130,000 iterations and a burn-in period equal to 60,000 iterations. Also, a thinning value of 13 was used to reduce the auto correlation in chains and hence 10,000 thinned steps were generated in each iteration.

3 Results

The Ivory Billed Woodpecker (IBW) *(Campephilus principalis)*, is one of the largest woodpeckers in the world but may have recently gone extinct. In the past decade several sightings of the IBW were reported but with uncertain validity, as it was impossible to obtain a clear photograph or other conclusive evidence of the bird (Collins 2017). A highly controversial uncertain sighting was recorded in 2004, and it was then argued that the IBW had been rediscovered. But whether the sighting was from the IBW or the Pileated Woodpecker *(Dryocopus pileatus)* is still open to debate (Sibley *et al.* 2007).

We proceed to analyse the sighting record data of the IBW provided in (Elphick *et al.* 2010) (see Supplementary Material (S1)), which gives 68 sightings throughout the period 1897 to 2009. Each of these sightings was classified into one of three different sighting classes.

1. Physical Evidence (PE) - e.g., museum specimens, but also uncontroversial photographs, video, and sound recordings.

2. Independent Expert Opinion (IEO) - evidence that experts deemed sufficiently documented to confirm the record.

3. Controversial sightings (CS) - sightings judged to lack firm evidence including any sighting for which there is published disagreement between experts.

Following Solow & Beet (2014), we consider only sightings belonging to the Physical Evidence (PE) class as certain while all other evidence as uncertain.

Model 1 - Certain sightings only

We begin by analysing the IBW data with the certain sighting only model (Model 1), and therefore initially dismiss the uncertain sightings IEO and CS from the sighting record, and analyse only the certain PE sightings. This requires working with the likelihood in Equation 3 and the prior distributions defined above. Then the posterior distribution of τ_E so obtained is summarized in Figure 1 and the 95% HDI for the posterior extinction year is given in Table 2. According to Table 2, the median extinction year is 1940 with a 95% upper bound in 1944. Also the posterior probability that extinction occurred during the observation period is equal to one, which gives overwhelming support that extinction occurred during the observation period. Based on these findings we can infer that the IBW went extinct within a few years after the last certain (i.e PE) sighting in 1939.

Table 2: Summary of the posterior distribution of τ_E using certain sightings only.

	95% HDI Low	median	95% HDI High
$\tau_E \mid S$	1940	1940	1944

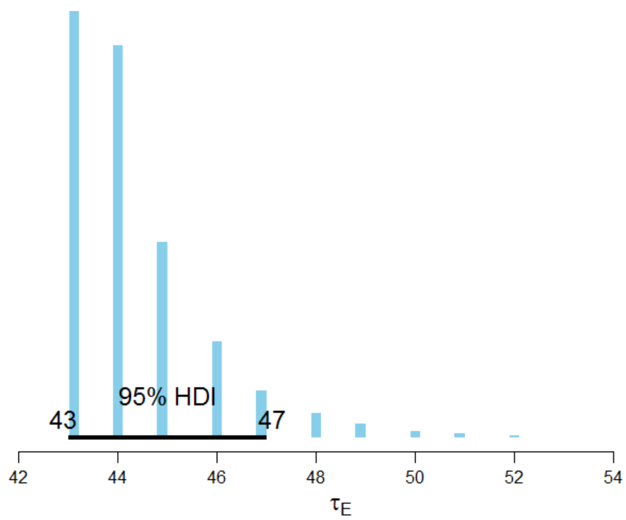
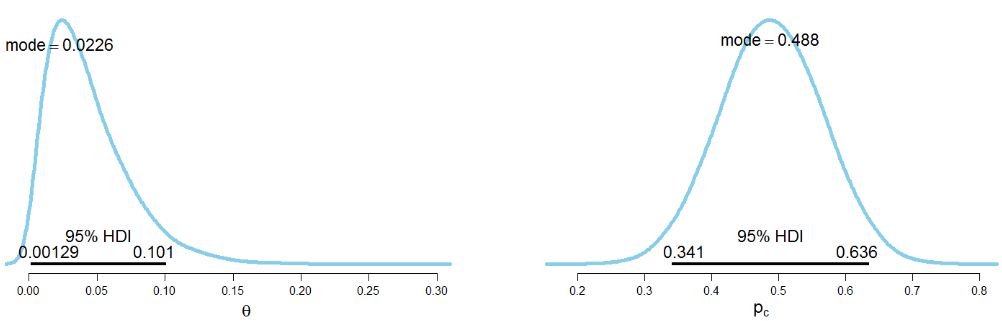


Figure 1: Posterior distribution plot of τ_E for the IBW for Model 1, where $\tau_E = 0$ refers to the year 1897. Black solid line above x-axis shows the 95% HDI for the posterior distribution.



(a) (b)

Figure 2: Posterior distribution plots for the model parameters for Model 1 excluding τ_E. Black solid line above x-axis shows the 95% HDI for the posterior distribution. (a) Posterior distribution of θ. (b) Posterior distribution of p_c.

The posterior distributions of model parameters θ and p_c are shown in Figure 2. Recall that a non-informative prior (i.e uniform distribution) was used for all these parameters. As per Figure 2a, the posterior estimate for the yearly extinction probability θ for the IBW is θ=0.02. Also, according to Figure 2b, the posterior estimate of p_c, which is the probability for recording a certain sighting is p_c=0.5, with a 95% HDI between 0.3 and 0.6.

Model 2 - Certain sightings and uncertain sightings

In Model 2, we follow Solow & Beet (2014) and assume that all PE sightings are certain, and all other sighting evidence (i.e IEO and CS) uncertain. We thus use the likelihood in Equation 9. The posterior distribution of τ_E is plotted in Figure 3 and the 95% HDI for the posterior extinction year is given in Table 3. According to Table 3, the median extinction year is 1940 with a 95% upper bound in 1945. Similar to Model 1, we can infer that the IBW went extinct within a few years of the last certain (i.e PE) sighting in 1939. Our findings contradict the results from a recent paper which predicts the extinction year for IBW to be much closer to the sighting end point in 2009 (Brook *et al.* 2019). Interestingly, the inference made concerning τ_E under Model 1 and Model 2 seems almost identical. Hence for the IBW sighting record, the inclusion of uncertain sightings does not affect the conclusion of the model, although this property is not always guaranteed (see Section 4).

Table 3: Summary of the posterior distribution of τ_E

	95% HDI Low	median	95% HDI High
$\tau_E \mid S$	1940	1940	1945

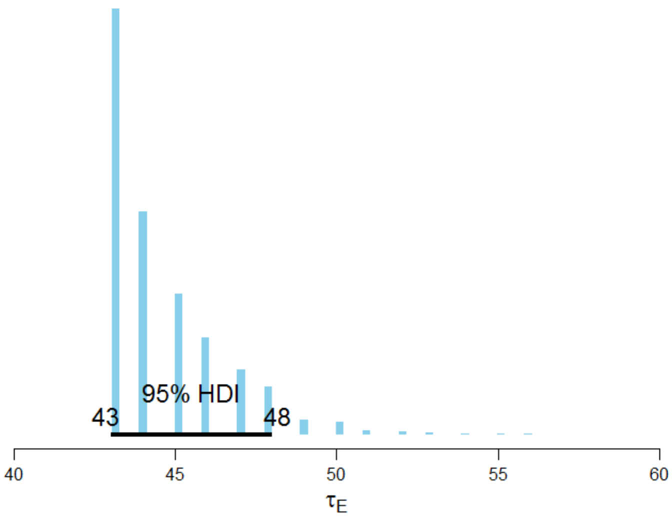


Figure 3: Posterior distribution plot of τ_E for the IBW for Model 2, where $\tau_E = 0$ refers to the year 1897. Black solid line above x-axis shows the 95% HDI for the posterior distribution.

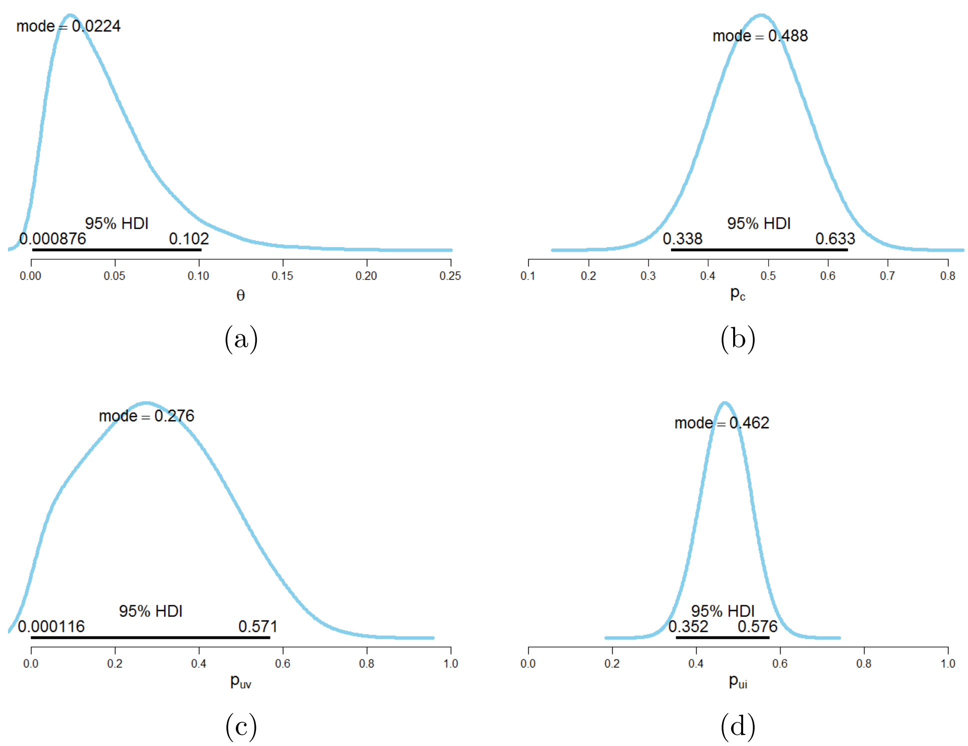


Figure 4: Posterior distribution plots for the model parameters for Model 2 excluding τ_E. Black solid line above x-axis shows the 95% HDI for the posterior distribution. (a) Posterior distribution of θ. (b) Posterior distribution of p_c. (c) Posterior distribution of p_{uv}. (d) Posterior distribution of p_{ui}.

The posterior distributions of other model parameters i.e θ, p_c, p_{ui} and p_{uv}, are shown in Figure 4. By comparing Figure 4c with Figure 4d it is noticeable that the value of the mode it is clear that there is a higher chance of observing a invalid uncertain sighting rather than a valid uncertain sighting. Also, the variability in the invalid uncertain probability is much less compared to the variability of the valid uncertain probability. Both Model 1 and Model 2 produce similar posterior distributions for for the yearly extinction probability θ and pc.

Treating uncertain sightings as certain

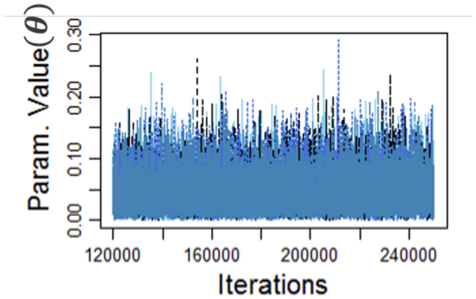
As an experiment, we now analyse the IBW data, treating all sightings (PE, IEO and CS) as certain sightings and just making use of Model 1. Under this assumption, the last certain sighting c_n is equal to the last (previously uncertain) sighting in 2007 and the total number of certain sightings is now equal to $n_c + N_u$. Based on these new inputs, it was found that the posterior estimate (median) for the extinction year for the IBW is increased to the year 2080 $\tau_E = 2080$, which is completely different to our previous results, and would suggest that the IBW is extant, if there might be reason to believe that the CS and IEO data were actually certain.

Diagnostic tests for MCMC Samples

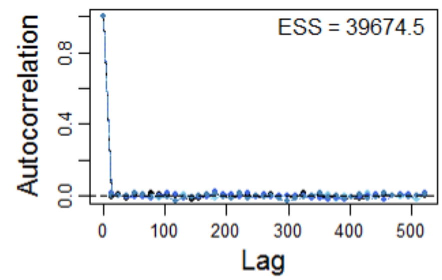
When using a Computational Bayesian approach it is important to carry out diagnostics checks to examine whether the quality of the MCMC chains are sufficient to provide an accurate approximation of the target distribution. In practice, the MCMC chains are often assessed through visual inspection of the trace plot, auto-correlation plot, shrink factor plot and marginal density plot. Addition to these visual inspections there are some numerical checks such as the effective sample size (ESS) and Monte Carlo standard error (MCSE) which are used to measure the accuracy of the chains. A full discussion on these tools can be found in (Kruschke 2014). Figure 5 illustrates these diagnostic checks for the parameter

θ using the IBW sightings.

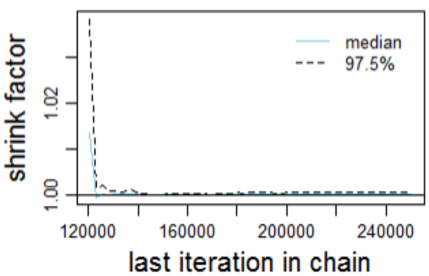
The trace plot in Figure 5a displays the values of the parameter θ (yearly extinction probability), during the run-time of the chain. This plot is used to identify any signs of irregular orphaned chains that might arise in some unusual regions of the parameter space. The plot given in Figure 5a indicates overlapping chains suggesting no orphaned chains. The marginal density plot of θ (see Figure 5d) is a smoothed histogram of the values in the trace-plot. This plot is used to identify if all the chains suitably represent the posterior distribution. The density plot also indicates overlapping chains, which suggest good representativeness of the posterior distribution. The auto-correlation plot given in Figure 5b indicates a zero auto-correlation between the chain values, which means that the values in a chain change rapidly for each and every step. As such, the chains are less clumpy and provide reasonably independent samples from the parameter distribution indicating that there are no problems. Inspection of convergence can also be checked numerically through the shrink factor, shown here in Figure 5c. A shrink factor above 1.1 indicates concerns on the convergence of the chains (Kruschke 2014), something that is not an issue in this example.



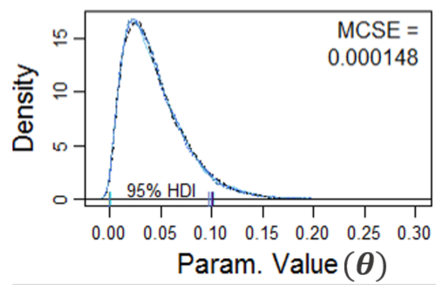
(a) Trace plot for θ



(b) Auto-correlation plot for θ



(c) Shrink factor plot for θ



(d) Marginal density plot for θ

Figure 5: Illustration of MCMC Diagnostics. The trace plot, auto-correlation plot, shrink factor plot and the marginal density plot outputted by JAGS. These plots are used to check if the chains are well mixed and suitably represent the posterior distribution. Analysis is based on data for the IBW (see text).

The density plot in Figure 5d displays the estimated 95% highest density interval (HDI) for each chain. The 95% HDI is a Bayesian credible interval, and values inside this interval have a total probability of 0.95. Because of the uncertainty in the parameter, HDI intervals for each chain will slightly differ from each other. The MCSE indicates the estimated standard deviation of the sample mean in the chain and an ESS value of at least 10,000 is desirable to have a reasonably accurate and stable estimate of the limits of the 95% HDI. As the ESS value for θ is around $40,000 (> 10,000)$ (see Figure 5b), the estimates for θ will be stable and accurate.

With the aid of Figure 5, we demonstrated how the MCMC chains generated for θ under Model 1 are sufficient to provide an accurate approximation for the target distribution. Similar diagnostic checks were carried out for all the model results presented in this paper

(Model 1/ Model 2) and no indication of any problem for any parameter (e.g. τ_E, p_c, p_{uv} etc.) was observed. All of these Diagnostic figures are give in Supplementary Material S2 and S3.

4 Sensitivity Analysis

In the previous section, we found that the inclusion of uncertain sightings changed the results of the IBW analysis very little compared to a model which omits them. Hence, it is important to see if this is a special case, or whether the uncertain sightings are generally non informative. To asses this we consider the three artificially generated sighting records shown in Figure 6. All thee time series have the same certain sighting history where it follows at a constant rate for the first 24 years of the 100 year observation record. While the first scenario has only certain sighting the second and the third includes uncertain sightings with different rates for the first 69 years.

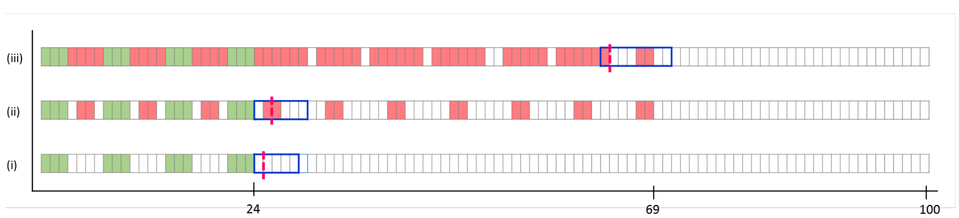


Figure 6: Posterior median extinction date and its 95% HDI for three artificially generated sighting records between 0 and 100. The cells shaded in green represents certain sightings while the red shades represent uncertain sightings. Also, the cells without any shade indicates no sightings. For each of the sighting record the posterior median extinction date is indicated from a pink dashed line and the 95% HDI interval in the blue region.

According to Figure 6 it is clear that the first two scenarios resulted in a median extinction date (shown by the pink dashed line) closer to the last certain sighting in year 24, while the third is father away and closer to the uncertain sightings change-point in year 64. A change-point can be defined as a point in time when the probability distribution of a sequence of sightings differs before and after the point. As per this definition the last certain sighting can also be considered as a change-point. Theoretically, extinction should happen closer to one of these change points as the rates of sightings (i.e certain/ uncertain) will fall following extinction. When there is only certain sightings there is only one change-point and hence extinction is much likely to happen after the last certain sighting. Also, when the uncertain sightings continue at a high rate after the last certain sighting and then fall to a low rate prior to the end of the observation period the uncertain sightings become more informative (see scenario (iii)). But when the uncertain sightings occur at a low constant rate (scenario (ii)) the change-point from uncertain sightings are not significant compared to the last certain sighting. Hence the result from scenario (ii) does not differ significantly from the scenario (i).

5 Discussion

In this study we present a Bayesian hierarchical approach to obtain the posterior distribution for τ_E (the date of the first year following extinction) and to calculate the posterior probability that the species is extinct by the end point of the sighting record data. Our general model is intended for sighting records that contain both certain and uncertain sightings. In order to obtain the posterior distribution for τ_E, we use Markov Chain Monte Carlo (MCMC) sampling techniques implemented with JAGS in R. As a case-study, we infer the extinction time distribution of τ_E for the IBW from historical sighting records.

In 2005, the IBW, which was thought to be extinct, received considerable attention after the announcement of its rediscovery in continental North America in the prestigious journal *Science* (Fitzpatrick *et al.* 2005). This announcement was based on a video clip analysis, which captured the species for a total of four seconds in 2004. However, the video had a number of imperfections, since images were blurred and pixelated owing to rapid motion, slow shutter speed, video interlacing artifacts, and the bird's distance beyond the video camera's focal plane (Fitzpatrick *et al.* 2005). Soon after the claim, Sibley et al. (Sibley *et al.* 2007) concluded that the evidence strongly suggests that the bird in the video was a normal pileated Woodpecker rather than an IBW, reigniting the controversy of whether the IBW was extinct or extant. Recent work has shown how drone technology may be used to find the IBW (Collins 2018) and resolve this controversy.

Our paper investigates another approach: using Bayesian statistical models to investigate whether the IBW is extinct. Two modelling approaches were presented. The first only dealt with historical records containing only certain sightings, while the second considered records that contain certain and uncertain sightings. We applied both approaches to the sighting data of the IBW assigning uniform priors to all model parameters. The null hypothesis that the IBW is extant by 2009 was rejected under both the certain sighting model (i.e Model 1) and the combined certain/uncertain sighting model (i.e. Model 2). Thus our statistical analysis suggests that the IBW went extinct in the 1940's, even when taking into account the uncertain sighting in 2006. Similar to our recent paper, the analysis highlighted the important role of the last certain sighting, especially when uncertain sightings are of low quality.

Through an artificially generated sighting records it was shown that the extinction is most likely to happen either near to the last certain sighting or to the point where the uncertain sightings fall to lower rate. These two time points can also be refereed as a change-point. Extinction is highly likely to occur at a change-point because the rate of sightings falls following extinction. There is only one change-point when we consider a certain sighting only scenario and that is the last certain sighting. In this case, the relatively high rate of certain sightings prior to the last certain sighting and their absence after that makes the last certain sighting to be a change-point. But when there is both certain and uncertain sightings then the significant change-point can occur from either sighting types. For example the uncertain sightings becomes informative about extinction in a situation in which uncertain sightings continue at a high rate after the last certain sighting and then fall to a low rate prior to the end of the observation period. In this situation the change-point is the point of time where the uncertain sighting rate changes. Interestingly the findings from our previous paper (Kodikara *et al.* 2018) also aligns with these findings where we showed that the two models developed in Solow and Beet (Solow & Beet 2014) were sensitive to different points. While their first model was sensitive to the last uncertain sighting, the second was sensitive to the last certain sighting. Hence extinction problem can be seen as a change-point analysis but this change-point will be dependent on the model assumptions. The findings from this paper can be used in a powerful manner in exploring extinction problems.

Data accessibility. The data-set supporting this article have been provided in the Supplementary Material (S1).

Competing interests. We declare we have no competing interests.

Authors' contributions. L.S, H.D. and Y.W. secured funding. S.K. and A.S. contributed to the concept and designed the article content while consulting other authors. S.K conducted the analyses and wrote the initial draft of the manuscript. All authors revised the manuscript and gave final approval for publication.

Funding. This work was funded by Australian Research Council[Grant number: DP 150102472].

References

Akçakaya, H., Keith, D.A., Burgman, M., Butchart, S.H., 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.

Alroy, J. (2014) A simple bayesian method of inferring extinction. *Paleobiology*, **40**, 584–607.

Alroy, J. (2016) On a conservative bayesian method of inferring extinction. *Paleobiology*, **42**, 670–679.

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

Battarbee, R.W. (2014) The rediscovery of the aldabra banded snail, rhachistia aldabrae.

Brook, B.W., Buettel, J.C. & Jarić, I. (2019) A fast re-sampling method for using reliability ratings of sightings with extinction-date estimators. *Ecology*, p. e02787.

Collins, M.D. (2017) Video evidence and other information relevant to the conservation of the ivory-billed woodpecker (campephilus principalis). *Heliyon*, **3**, e00230.

Collins, M.D. (2018) Using a drone to search for the ivory-billed woodpecker (campephilus principalis). *Drones*, **2**, 11.

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

Fader, P.S., Hardie, B.G. & Shang, J. (2010) Customer-base analysis in a discrete-time noncontractual setting. *Marketing Science*, **29**, 1086–1108.

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., Wrege, P.H. *et al.* (2005) Ivory-billed woodpecker (campephilus principalis) persists in continental north america. *Science*, **308**, 1460–1462.

Gerlach, J. (2007) Short-term climate change and the extinction of the snail rhachistia aldabrae (gastropoda: Pulmonata). *Biology Letters*, **3**, 581–585.

Jarić, I. & Roberts, D.L. (2014) Accounting for observation reliability when inferring extinction based on sighting records. *Biodiversity and conservation*, **23**, 2801–2815.

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.

Kodikara, S., Demirhan, H. & Stone, L. (2018) Inferring about the extinction of a species using certain and uncertain sightings. *Journal of theoretical biology*, **442**, 98–109.

Kodikara, S., Demirhan, H., Wang, Y., Solow, A. & Stone, L. (2020) Inferring extinction year using a bayesian approach. *Methods in Ecology and Evolution*, **11**, 964–973.

Kruschke, J. (2014) *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan.* Academic Press.

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

Lee, T.E., Bowman, C. & Roberts, D.L. (2017) Are extinction opinions extinct? *PeerJ*, **5**, e3663.

Lee, T.E., McCarthy, M.A., Wintle, B.A., Bode, M., Roberts, D.L. & Burgman, M.A. (2014) Inferring extinctions from sighting records of variable reliability. *Journal of Applied Ecology*, **51**, 251–258.

Marshall, C.R. (1990) Confidence intervals on stratigraphic ranges. *Paleobiology*, **16**, 1–10.

Pimm, S.L., Jenkins, C.N., Abell, R., Brooks, T.M., Gittleman, J.L., Joppa, L.N., Raven, P.H., Roberts, C.M. & Sexton, J.O. (2014) The biodiversity of species and their rates of extinction, distribution, and protection. *Science*, **344**, 1246752.

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., Elphick, C.S. & Reed, J.M. (2010) Identifying anomalous reports of putatively extinct species and why it matters. *Conservation Biology*, **24**, 189–196.

Sibley, D.A., Bevier, L.R., Patten, M.A. & Elphick, C.S. (2007) Ivory-billed or pileated woodpecker? *Science*, **315**, 1495–1496.

Solow, A., Smith, W., Burgman, M., Rout, T., Wintle, B. & Roberts, D. (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. (2016a) On bayesian inference about extinction. *Proceedings of the National Academy of Sciences*, **113**, E1132–E1132.

Solow, A.R. (2016b) A simple bayesian method of inferring extinction: comment. *Ecology*, **97**, 796–798.

Solow, A.R. & Beet, A.R. (2014) On uncertain sightings and inference about extinction. *Conservation biology*, **28**, 1119–1123.

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

Thompson, C., Lee, T., Stone, L., McCarthy, M. & Burgman, M. (2013) Inferring extinction risks from sighting records. *Journal of Theoretical Biology*, **338**, 16–22.

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.