



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Wintle, BA;Walshe, TV;Parris, KM;Mccarthy, MA

Title:

Designing occupancy surveys and interpreting non-detection when observations are imperfect

Date:

2012-04-01

Citation:

Wintle, B. A., Walshe, T. V., Parris, K. M. & Mccarthy, M. A. (2012). Designing occupancy surveys and interpreting non-detection when observations are imperfect. *Diversity and Distributions*, 18 (4), pp.417-424. <https://doi.org/10.1111/j.1472-4642.2011.00874.x>.

Persistent Link:

<https://hdl.handle.net/11343/217111>

Designing occupancy surveys and interpreting non-detection when observations are imperfect

Brendan A. Wintle*¹, Terry V. Walshe¹, Kirsten M. Parris¹ and Michael A. McCarthy¹

¹ School of Botany, University of Melbourne, Victoria, Australia, 3010

*Corresponding author. Email: brendanw@unimelb.edu.au

ph) +61 3 8344 3306

Terry Walshe: twalshe@unimelb.edu.au

Kirsten Parris: kparris@unimelb.edu.au

Michael McCarthy: mamcca@unimelb.edu.au

Running title: Interpreting non-detections in occupancy surveys

Word count:

Abstract: 280

Main text: 3627

No. References: 34

No. Tables: 0

No. Figures: 3

1 **ABSTRACT**

2 **Aim** Conservation practitioners use biological surveys to ascertain whether or not a site is
3 occupied by a particular species. Widely-used statistical methods estimate the probability that a
4 species will be detected in a survey of an *occupied* site. However, these estimates of detection
5 probability are alone not sufficient to calculate the probability that a species is present given that it
6 was not detected. The aim of this paper is to demonstrate methods for correctly calculating (i) the
7 probability a species occupies a site given one or more non-detections, and (ii) the number of
8 sequential non-detections necessary to assert, with a pre-specified confidence, that a species is absent
9 from a site.

10 **Location** Occupancy data for a tree frog in eastern Australia serve to illustrate methods that
11 may be applied anywhere species' occupancy data are used and detection probabilities are less than 1.

12 **Methods** Building on Bayesian expressions for the probability that a site is occupied by a
13 species when it is not detected, and the number of non-detections necessary to assert absence with a
14 pre-specified confidence, we estimate occupancy probabilities across tree frog survey locations,
15 drawing on information about where and when the species was detected during surveys.

16 **Results** We show that the number of sequential non-detections necessary to assert that a
17 species is absent increases non-linearly with the prior probability of occupancy, the probability of
18 detection if present, and the desired level of confidence about absence.

19 **Main conclusions** If used more widely, the Bayesian analytical approaches illustrated here
20 would improve collection and interpretation of biological survey data; providing a coherent way to
21 incorporate detection probability estimates in the design of minimum survey requirements for
22 monitoring, impact assessment and distribution modelling.

23 **Key-words:** Bayes' theorem, detectability, survey effort, monitoring, species distribution model

24 **(A) Introduction**

25 Species site occupancy data underpin many of the analyses undertaken in conservation
26 biogeography. Of eight 'prominent areas of research in conservation biogeography' identified by
27 Richardson and Whittaker (2010), five are fundamentally reliant on occupancy data; (i)
28 understanding processes such as extinction, persistence, range expansion, dispersal; (ii) inventory
29 and mapping; (iii) species distribution modelling; (iv) characterizing biotas, including species–area
30 relationships, and (v) conservation planning. Each of these activities is, to some extent compromised
31 by uncertainty arising from imperfect detection of species during biological surveys. A range of
32 statistical methods exist to model imperfect detection of species during occupancy surveys, estimate
33 species' detection probabilities, identify conditions most conducive to detection, and control for
34 imperfect detection in statistical inference (McArdle, 1990; Boulinier *et al.*, 1998; MacKenzie *et al.*,
35 2002; MacKenzie *et al.*, 2003; Tyre *et al.*, 2003; Wintle *et al.*, 2004; Royle & Link, 2006). These
36 statistical approaches have been primarily used to estimate detection probabilities under various
37 survey conditions (Bailey *et al.*, 2004; Wintle *et al.*, 2005), to analyse temporal trends in habitat
38 occupancy (MacKenzie *et al.*, 2002; MacKenzie *et al.*, 2003; Field *et al.*, 2005), to condition species
39 richness estimates (Dorazio *et al.*, 2006 ; Kéry *et al.*, 2009), and to remove false negative
40 observation bias from estimates of species distribution model coefficients (e.g. Tyre *et al.*, 2003).

41
42 Arguably, the most common application of detectability estimates is in interpreting observation data
43 to determine whether or not a species is, in fact, present at a given site when not detected.

44 Environmental impact assessments utilize these kinds of data to inform decisions about whether or
45 not destruction or development of potential habitats should be allowed to proceed, at the risk of
46 losing endangered species that have not been detected on the site (Garrard *et al.*, 2008). Declaring
47 eradication of a weed or disease depends on the probability that there are unobserved breeding
48 individuals (Regan *et al.*, 2006; Rout *et al.*, 2010). Quarantine operations must assess the probability

49 that an unwanted pest is in fact present in a shipping container, given that it was not detected using a
50 particular search strategy (Burgman *et al.*, 2010). In all such cases, it is important to correctly
51 interpret non-detection data so that decisions can be based on coherent estimates of the probability
52 that a species is truly absent, or alternatively that the species is present but not detected.

53

54 It is tempting to imagine that a good estimate of the detection probability; the probability that a
55 species would be detected *if* it is present, would be enough to estimate the probability it is present
56 given that it was not detected in a given number of survey visits. Unfortunately, it is not enough to
57 know the probability of detection conditional on presence if the aim is to determine the probability of
58 presence given non-detections. To illustrate, let's say the chance of detecting a hypothetical rare
59 species, *if in fact it is present*, is 50% in any one survey and that six independent surveys at a site fail
60 to detect it. There is a probability of $(0.5)^6 = 0.016$ that all six surveys will fail to detect the species *if*
61 it is present. There is a tendency to confuse this, the probability that species is not detected given that
62 it is present, with the probability that it is present given that it was not detected (e.g., Pellet &
63 Schmidt, 2005; Jackson *et al.*, 2006; Olea & Mateo-Tomas, 2011). This common logical error is
64 known as the 'inverse fallacy' or 'base-rate fallacy' (Bar-Hillel, 1980; Koehler, 1996; Villejoubert &
65 Mandel, 2002) and amounts to mistakenly accepting that $\Pr(A|B) = \Pr(B|A)$. This result also impacts
66 on the design of minimum survey effort requirements for detecting species. If one wishes to calculate
67 the number of sequential non-detections necessary to assert, with a pre-specified confidence, that a
68 species is truly absent, it is not sufficient to consider only the detection probability. One must also
69 consider the expected prevalence of positive observations (expected rate of occupancy in a sample).
70 This quantity is equivalent to the prior probability of occupancy in a Bayesian analysis.

71

72 Bayesian approaches have been applied to modelling imperfect detection data (e.g., Wintle *et al.*,
73 2005; McCarthy, 2007; Royle *et al.*, 2007; Garrard *et al.*, 2008; Royle & Dorazio, 2008; Burgman *et*

74 *al.*, 2010). The aim of this paper is to demonstrate Bayesian approaches to collecting and interpreting
75 observation data of the kind described above. We provide mathematical expressions and computer
76 code to (i) estimate the probability that a species occupies a given site after one or more non-
77 detections have occurred at that site; (ii) estimate the number of sequential non-detections necessary
78 to assert with a pre-specified degree of confidence that a species is truly absent from the site; and (iii)
79 generalize these to estimate occupancy probabilities at multiple sites, drawing on information about
80 the sorts of places the species has and has not been located over all the sites in a multi-site survey.
81 We demonstrate the application of these methods using a case study based on tree frog survey data
82 from sub-tropical eastern Australia.

83

84 **(A) Methods**

85 (B) Model

86 The correct logical structure of the problem is more accessible if we draw it as a logic tree (Fig. 1),
87 and use frequencies instead of probabilities (Gigerenzer & Hoffrage, 1995). Remember our
88 hypothetical species that is detected on average 50% of the time during individual surveys to
89 occupied sites. There is a probability of $(0.5)^6 = 0.016$ that it would remain undetected in 6 visits to a
90 site if it is present there. Let's say that past records indicate the species was present at about one in
91 four sites having comparable habitat. If we imagine 1000 such sites, the species is expected to be
92 present at 250. Of those, six repeat surveys at each site will detect the species at $(1 - 0.016) \times 250 =$
93 246 sites. If the species is not detected, it's either a false absence (4/1000) or a true absence
94 (750/1000). The chance the species is actually present despite six surveys reporting absence is $4/(4 +$
95 750) = 0.005. Note that this probability is conditioned by the first branch of the logic tree; the
96 expected true rate of occupancy (or the prior belief the species is present). If our prior belief is that
97 the species will be present at about three in four sites of comparable habitat, the corresponding
98 posterior probability of occupancy is $12/(12 + 250) = 0.046$, almost an order of magnitude greater.

99 The simple calculations illustrated in the logic tree are equivalent to the solution based on Bayes'
100 Theorem as we now show.

101

102 Let p' be a prior probability that a species that is in fact present will be detected in any single survey
103 of a fixed effort at a single site. The likelihood of a single non-detection if the species is in fact
104 present is $1 - p'$. If Ψ' is the prior probability that the species occupies that site, then Bayes' theorem
105 gives the posterior probability of the site being occupied given that it was not detected in a single
106 survey (Wintle *et al.*, 2005):

107

$$108 \quad p = \Psi'(1-p') / (\Psi'(1-p') + (1-\Psi')) \quad (\text{eq. 1})$$

109

110 The posterior probability of absence is then simply $1 - \Psi$.

111

112 Bayes' theorem for the posterior probabilities of presence and absence can be generalised to the case
113 where there are n sequential survey visits to a site in which the species was not detected. In this case,
114 and given independence of detections among visits, the likelihood of observing a sequence of n non-
115 detections at a site that is occupied is $(1-p')^n$. The posterior probability that the site is occupied (Ψ) is
116 then (Wintle *et al.*, 2005):

117

$$118 \quad \Psi = \Psi'(1-p')^n / (\Psi'(1-p')^n + (1-\Psi')) \quad (\text{eq. 2})$$

119

120 Note that the model ignores the possibility of false presences arising from misidentification of
121 species, though it may be extend to do so (Bar-Hillel, 1980; Royle & Link, 2006).

122

123 Rearranging equation 2 to solve for n provides the number of sequential non-detections (n^*)
124 necessary to achieve a particular posterior probability of absence from the site ($1-\Psi$). This takes into
125 account a prior belief about detectability of the species (p') and the prior (before collection of data)
126 probability that the species is present (Ψ') (Wintle *et al.*, 2005):

127

$$128 \quad n^* \Rightarrow \frac{\log\left(\frac{\Psi}{1-\Psi}\right) - \log\left(\frac{\Psi'}{1-\Psi'}\right)}{\log(1-p')}, \quad (\text{eq. 3})$$

129

130 Plotting equation 3 highlights the non-linear interaction between the occupancy prior (Ψ') and the
131 detection probability (p') in determining the number of sequential non-detections necessary to
132 achieve a pre-specified posterior level of confidence in the inference of true absence (Fig. 2).

133

134 Ecologists often collect repeat survey occupancy data across numerous sites in a study area, either to
135 estimate an overall rate of habitat occupancy in the case of monitoring applications (Field *et al.*,
136 2005), or to statistically infer species-environment relationships in the form of species distribution
137 models (Gu & Swihart, 2004). In both cases, it is important to account for imperfect detectability to
138 avoid biased inference. It is therefore useful to generalize equation 2 to estimate occupancy
139 probabilities using multi-site, multi-visit survey data, taking into account site- and visit-level
140 variation in detectability and probability of occupancy due to environmental conditions and the
141 observation process.

142

143 At a site i occupied by a species of interest, the likelihood of observing the species in the j th visit to
144 the site is p_{ij} and the likelihood of failing to observe the species is $1 - p_{ij}$. Site- and visit-level
145 detection probabilities may vary due to environmental influences on detectability such as vegetation
146 density or visit-level factors such as ambient weather conditions (Wintle *et al.*, 2005), or the survey
147 method used on a given visit to a site (Parris *et al.*, 1999). Let Y_i represent a vector (sequence) of

148 observations of length v_i (the number of visits to the site). Each element of the vector y_{ij} may take the
 149 value of 1 if the species was observed in the j th visit and 0 if the species was not observed in the j th
 150 visit. The likelihood of a given detection history (Y_i) over v visits to a site i is therefore (MacKenzie
 151 *et al.*, 2002):

$$152 \quad L(Y_i | \Psi_i, p_{ij}) = \Psi_i \prod_{j=1}^{v_i} p_{ij}^{y_{ij}} (1 - p_{ij})^{1-y_{ij}} , \quad \sum_{j=1}^{v_i} y_{ij} > 0 \quad (\text{eq. 4})$$

$$153 \quad L(Y_i | \Psi_i, p_{ij}) = \Psi_i \prod_{j=1}^{v_i} (1 - p_{ij}) + (1 - \Psi_i) , \quad \sum_{j=1}^{v_i} y_{ij} = 0 \quad (\text{eq. 5})$$

154
 155
 156 Having defined the likelihoods for multi-site, multi-visit data with respect to the site occupancy
 157 probabilities (Ψ_i) and the detection probabilities (p_{ij}), a Bayesian approach to generating posterior
 158 estimates of Ψ_i and p_{ij} requires a prior for each. If there is a reasonable expectation that
 159 environmental and observation process variables are likely to influence the p_{ij} and Ψ_i , such that they
 160 may vary over different sites or visits to those sites, it makes sense to model these probabilities as a
 161 function of environmental variables using an appropriate regression method (McCullagh & Nelder,
 162 1989):

$$163 \quad \text{logit}(p_{ij}) = \alpha + \sum_{k=1}^K \beta_k X_{ik} + \sum_{m=1}^M \gamma_m Z_{ijm} , \quad (\text{eq. 6})$$

$$164 \quad \text{logit}(\Psi_i) = \sigma + \sum_{k=1}^K \delta_k Y_{ik} , \quad (\text{eq. 7})$$

165
 166
 167
 168 where the α , σ , β_k , γ_m and δ_k are regression coefficients indicating the strength of the influence of
 169 environmental variables Y_k , X_k , and Z_m on occupancy and detection probabilities. In this case, the X_k
 170 and Y_k vary across sites, while the Z_m vary across both sites (indexed by i) and visits (indexed by j),
 171 and could be comprised of environmental, weather and observation variables such as observer
 172 experience or observation method.

173

174 Prior probability distributions are required for all of the α , σ , β_k , and γ_m regression coefficients. In
175 this case there is little basis for strong belief in any prior, so a reasonable choice would be
176 uninformative normal distributions with a mean equal to 0 and large variance. A full Bayesian
177 analysis of multi-visit, multi-site observation data using this model in a Bayesian modelling package
178 such as OpenBUGS (Lunn *et al.*, 2000) yields posterior estimates of the strength of influences of
179 environmental, weather and observation processes on both species occupancy and species
180 detectability. By substituting Ψ'_i and p'_{ij} into Eq.3, it is then possible to estimate the required survey
181 effort (n^*), as a function of the values of site and survey conditions (*i.e.*, as a function of the X and
182 the Z from equations 6 and 7). Uncertainty about the Ψ'_i and p'_{ij} can be propagated through the
183 calculation of n^* using Bayesian software such as OpenBUGS (see Appendix S1 in supporting
184 information for all OpenBUGS code used in our analyses).

185

186 In the following section, we demonstrate the application of the models described above by analysing
187 multi-site, multi-visit survey data for the cascade tree frog (*Litoria pearsoniana*) in eastern Australia.

188

189 (B) Application Data

190 *Litoria pearsoniana* is a tree frog that breeds in forest streams in sub-tropical eastern Australia.
191 Surveys of 64 sites throughout its range in south-east Queensland and north-east New South Wales
192 were conducted over an area of approximately 14,000 km² between 1995-1999 (Parris, 2001). Two
193 survey methods were employed; nocturnal searches and automatic tape recording of advertisement
194 calls. The data comprise a record of the detection or non-detection of the species on each survey
195 night at each survey site. The only visit-level variable considered that could have influenced the
196 probability of detection in each survey is the type of survey method used (search versus tape
197 recording). Variables thought most likely to influence the probability of site occupancy by *L.*

198 *pearsoniana* were the catchment volume of the stream, indicating the permanence and volume of
199 stream flow, and the presence or absence of palms at the site, which indicates mesic or xeric
200 conditions in the riparian zone (Parris, 2001). Catchment volume was calculated as the mean annual
201 volume of rain that fell in the catchment upstream of the site and ranged from 114 to 102,000
202 gigalitres across survey locations.

203

204 (B) Application model

205 Data of Parris (2001) were re-modelled using the freeware Bayesian modelling package OpenBUGS
206 3.1.2 (Lunn *et al.*, 2000). The model set-up was identical to that described in equations 4-7. The
207 analysis of visit-level variation in detectability (p_{ij}) was simplified by having only a single
208 categorical variable (survey method) influencing p_{ij} . The influence of the explanatory variables
209 catchment volume (modelled as the natural log of catchment volume: $\ln CV$) and the presence of
210 palms (*palms*) on the probability of *L. pearsoniana* occupancy was modelled using logistic
211 regression (McCullagh & Nelder, 1989). A multiplicative interaction term for these two variables
212 was also included. Uncertainty about the strength of influence of the explanatory variables on *L.*
213 *pearsoniana* occupancy prior to data analysis was characterised using uninformative normal
214 distributions on regression coefficients with a mean of zero and standard deviation of 1000. Prior
215 uncertainty about the detectability of *L. pearsoniana* with the two survey methods was characterised
216 using uninformative uniform prior distributions between zero and one. Posterior distributions for i)
217 the regression model coefficients, ii) probabilities of presence over the observed range of the
218 explanatory variables, and iii) nightly detection probabilities for the two survey methods were
219 obtained from 50 000 Markov chain Monte-Carlo (MCMC) samples after discarding a 10 000
220 sample burn-in (Appendix S1).

221

222 (A) Results

223 Analysis confirmed a strong positive influence of catchment volume and a weak positive influence of
224 palms on the occurrence of *L. pearsoniana*, and a strong positive interaction between the two
225 variables (the effect of catchment volume is stronger in the presence of palms) (Appendix S1). At
226 the sites with the lowest catchment volumes in the study, the probability of them containing tree
227 frogs was slightly higher in the absence of palms (~0.1) compared with sites in which palms were
228 present (~ 0.05). However, when a site was situated in a medium or larger sized catchment, the
229 probability of tree frog occupancy was more than tripled at sites with palms compared to sites
230 without (0.2-0.3 small catchment versus 0.7-0.9 large catchments. The mean probability of detection
231 using nocturnal searches was estimated to be 0.56, which is substantially higher than the mean
232 detection probability arising from automatic recording of calls (0.35).

233

234 (B) Minimum survey effort calculations

235 By utilizing equation three in the OpenBUGS detectability and occupancy model of *L. pearsoniana*
236 (computer code in supplementary material) we were able to estimate the minimum survey effort
237 (number of repeat visits) necessary to achieve some pre-specified confidence in a conclusion that the
238 species was truly absent from a particular location under a range of environmental and detectability
239 conditions that may be encountered in future surveys (Fig. 3). It is apparent from Figure 3 that the
240 number of sequential non-detections necessary to be 95% sure that the species is absent from a given
241 site increases as the variables that positively influence probability of *L. pearsoniana* occupancy
242 increase. Under the most effective survey method (spotlighting streams), a 10-fold increase in the
243 volume of streams with palms leads to a 2-3-fold increase in the number of non-detections necessary
244 to be 95% certain that the species is, in fact, absent. This is because the prior probability that the
245 species occupies larger streams is substantially higher than that for smaller streams, necessitating a
246 greater weight of evidence (in the form of sequential non-detections) to provide the same level of
247 (posterior; after data) confidence that the species is absent. At the highest level of catchment volume

248 recorded in the study, in a site containing palms, approximately 9 sequential non-detections using
249 spotlighting surveys are required to be 95% sure the site is unoccupied, compared with the 18 non-
250 detections using tape recording that would be needed for the same level of confidence in absence
251 (Fig 3).

252

253 **(A) Discussion**

254 In the models described and demonstrated here, the role of a prior belief (in the form of a prior
255 probability of species occupancy) is central to a coherent interpretation of non-detections in survey
256 results. While some readers may feel uneasy about the use of prior probabilities (especially
257 subjective prior probabilities), failure to consider prior expectations, also known as 'base rates' or
258 expected prevalence (Koehler, 1996), is likely to lead to logical flaws in data interpretation,
259 including the 'inverse fallacy'. Utilizing previous studies or previous season's data to derive priors for
260 the expected rate of occupancy (or prevalence of positive observations) would generally be the
261 preferred means of estimating the prior probability of occupancy for those wishing to minimize
262 subjectivity.

263

264 However, in the design and analysis of field experiments, it is common to *implicitly* utilize prior
265 information. For example, if an ornithologist is searching for the northern spotted owl in a highly
266 productive, mature Douglas Fir forest in North America with a rich small mammal faunal
267 assemblage, they are likely to harbour a strong prior belief that the owl is present somewhere in the
268 area and may require a substantial number of non-detections to convince them otherwise. If the
269 search is being conducted in marginal habitat, a lesser effort may be intuitively employed.

270

271 The insights from models developed here emphasise the importance of *explicit* estimation and use of
272 prior beliefs. Estimates may be based directly on biological judgment, the predictions generated

273 from a habitat model, or simply the *unconditional rate of occupancy* (also known as expected
274 prevalence or the 'base rate') from previous surveys of the species (MacKenzie, 2005). An
275 uninformative prior probability of occupancy, $\Psi' = 0.5$ may be difficult to justify in many instances.
276 For example, consider a species that on the basis of historical records is estimated to be present at
277 10% of sites within a study region of variable habitat quality. If $p = 0.3$ and we wish to be 99%
278 confident of absence, then from equation 3, we require 13 sequential non-detections should we insist
279 on use of the uninformative prior, $\Psi' = 0.5$. If we use $\Psi' = 0.10$, then seven non-detections are
280 required. When surveying resources are scarce, use of an uninformative prior represents an
281 opportunity cost. A sophisticated approach that recognises opportunity costs would utilise biological
282 judgment to discern areas where the species is more (or less) likely to be present than the overall
283 10% estimate of prevalence.

284

285 Despite the fundamental importance of prevalence in conditioning estimates of species absence, we
286 could find no published examples in ecology where expected or previously observed prevalence
287 were explicitly incorporated in the design of a survey, let alone used to determine the required survey
288 effort. The advantage of our approach is that potentially implicit and subjective judgements are made
289 explicit, and the consequences of those judgements can be enumerated. Equation (3) makes clear that
290 decisions about necessary survey effort to determine the status of a species at a site depend on the
291 suitability of the site (Ψ'), the reliability of the survey (p), and the probability of occupancy required
292 when the survey fails to detect the species (Ψ). Scientific methods are available to estimate Ψ' and p ,
293 yet the required posterior probability of presence (Ψ) depends on social and political judgements that
294 reflect the costs of false absences. False absences in impact assessment for endangered species or
295 surveillance for invasive species might incur costs due to elevated risks of local or global extinction,
296 or of establishment and spread of a pest (Regan *et al.*, 2006). These costs need to be weighed against

297 the costs of additional survey effort. Our models support a framework for estimating the survey
298 effort that will lead to least overall cost.

299

300 The methods presented and illustrated here provide a basic toolkit for interpreting and dealing with
301 non-detections in biological surveys. There are a multitude of variations on the methods we describe
302 that will be necessary for interpreting occupancy data under survey designs and analytical constraints
303 that we have not addressed. For example, temporal dependence in detections would violate the
304 independence assumption necessary for using equation 3, in which case, correlations in detections
305 might need to be accounted for explicitly. However, adopting the general approach to data
306 interpretation and analysis presented here will increase the utility of existing methods for analysing
307 data under imperfect detection conditions. In particular, explicit consideration of prior beliefs and
308 analysis within a Bayesian analytical framework allows an interpretation of biological survey data
309 that is more intuitive and more useful for decision making.

310

311 **(A) Acknowledgements**

312 BAW was supported by an ARC Fellowship under DP0774288. BAW, MMC, KMP were supported
313 by funding from the National Environment Research Program Environmental Decisions Hub. TVW
314 was supported by the Australian Centre for Excellence in Risk Analysis. Georgia Garrard, Mark
315 Burgman and Libby Rumpff provided helpful references and comments.

316

317 **(A) References**

318 Bailey, L.L., Simons, T.R. & Pollock, J.H. (2004) Estimating site occupancy and species detection probability
319 parameters for terrestrial salamanders. *Ecological Applications*, **14**, 692-702.

320 Bar-Hillel, M. (1980) The base-rate fallacy in probability judgments. *Acta Psychologica*, **44**, 211 - 233.

321 Boulinier, T., Nichols, J.D., Sauer, J.R., Hines, J.E. & Pollock, K.H. (1998) Estimating species richness: the
322 importance of heterogeneity in species detectability. *Ecology*, **79**, 1018-1028.

323 Burgman, M.A., Wintle, B.A., Thompson, C.A., Moilanen, A., Runge, M.C. & Ben-Haim, Y. (2010)
324 Reconciling uncertain costs and benefits in Bayes nets for invasive species management *Risk analysis*,
325 **30**, 277-284.

326 Dorazio, R.M., Royle, J.A., Söderström, B. & Glimskär, A. (2006) Estimating species richness and
327 accumulation by modeling species occurrence and detectability. *Ecology*, **87**, 842-54.

328 Field, S.A., Tyre, A.J., Thorn, K.H., O'Conner, P.J. & Possingham, H.P. (2005) Improving the efficiency of
329 monitoring by estimating detectability: A case study of foxes (*Vulpes vulpes*) on the Eyre Peninsula,
330 South Australia. *Wildlife Research*, **32**, 252-258.

331 Garrard, G.E., Bekessy, S.A., McCarthy, M.A. & Wintle, B.A. (2008) When have we looked hard enough? A
332 novel method for setting minimum survey effort protocols for flora surveys. *Austral Ecology*, **33**, 986-
333 998.

334 Gigerenzer, G. & Hoffrage, U. (1995) How to improve Bayesian reasoning without instruction: frequency
335 format. *Psychological Review*, **102**, 684-704.

336 Gu, W. & Swihart, R.K. (2004) Absent or undetected? Effects of non-detection of species occurrence on
337 wildlife-habitat models. *Biological Conservation*, **116**, 195-203.

338 Jackson, J.T., Weckerly, F.W., Swannack, T.M. & Forstner, M.R.J. (2006) Inferring absence of Houston
339 Toads given imperfect dection probabilities. *Journal of Wildlife Management*, **70**, 1461-1463.

340 Kéry, M., Royle, J.A., Plattner, M. & Dorazio, R.M. (2009) Species richness and occupancy estimation in
341 communities subject to temporary emigration. *Ecology*, **90**, 1279-90.

342 Koehler, J.J. (1996) The base rate fallacy reconsidered: Descriptive, normative and methodological
343 challenges. *Behavioral & Brain Sciences*, **19**, 1-53.

344 Lunn, D.J., Thomas, A., Best, N. & Spiegelhalter, D. (2000) WinBUGS; a Bayesian modelling framework:
345 concepts, structure, and extensibility. *Statistics and Computing*, **10**, 325--337.

346 MacKenzie, D.I. (2005) What are the issues with presence-absence data for wildlife managers? *Journal of*
347 *Wildlife Management*, **69**, 849-860.

348 MacKenzie, D.I., Nichols, J.D., Hines, J.E., Knutson, M.G. & Franklin, A.B. (2003) Estimating site
349 occupancy, colonization and local extinction when a species is detected imperfectly. *Ecology*, **84**,
350 2200-2207.

351 MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A. & Langtimm, C.A. (2002) Estimating
352 site occupancy rates when detection probabilities are less than one. *Ecology*, **83**, 2248-2255.

353 McArdle, B.H. (1990) When are rare species not there? *Oikos*, **57**, 276-278.

354 McCarthy, M.A. (2007) *Bayesian methods for Ecology*. Cambridge University Press, Cambridge.

355 McCullagh, P. & Nelder, J.A. (1989) *Generalised linear models*, 2nd edn. Chapman and Hall, London,
356 England.

357 Olea, P.P. & Mateo-Tomas, P. (2011) Spatially explicit estimation of occupancy, detection probability and
358 survey effort need to inform conservation planning. *Diversity and distributions*, **online early**

359 Parris, K.M. (2001) Distribution, habitat requirements and conservation of the cascade treefrog (*Litoria*
360 *pearsoniana*, Anura: Hylidae). *Biological Conservation*, **99**, 285-292.

361 Parris, K.M., Norton, T.W. & Cunningham, R.B. (1999) A comparison of techniques for sampling amphibians
362 in the forests of south-east Queensland, Australia. *Herpetologica*, **55**, 271-283.

363 Pellet, J. & Schmidt, B.K. (2005) Monitoring distributions using call surveys: estimating site occupancy,
364 detection probabilities and inferring absence. *Biological Conservation*, **123**, 27–35.

365 Regan, T.J., McCarthy, M.A., Baxter, P.W.J., Panetta, F.D. & Possingham, H.P. (2006) Optimal eradication:
366 when to stop looking for an invasive plant. *Ecology Letters*, **9**, 759-766.

367 Richardson, D.M. & Whittaker, R.J. (2010) Conservation biogeography - foundations, concepts and
368 challenges. *Diversity and distributions*, **16**, 313–320.

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

371 Royle, J.A. & Link, W.A. (2006) Generalized site occupancy models allowing for false positive and false
372 negative errors. *Ecology* **87**, 835-841.

373 Royle, J.A. & Dorazio, R.M. (2008) *Hierarchical modeling and inference in ecology*. Academic Press,
374 Amsterdam.

375 Royle, J.A., M. Kéry, R. Gautier & Schmid., H. (2007) Hierarchical spatial models of abundance and
376 occurrence from imperfect survey data. *Ecological Monographs*, **77**, 465-481.

377 Swets, J.A. (1988) Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285-1293.

378 Tyre, A.J., Tenhumberg, B., Field, S.A., Possingham, H.P., Niejalke, D. & Parris, K. (2003) Improving
379 precision and reducing bias in biological surveys by estimating false negative error rates in presence-
380 absence data. *Ecological Applications*, **13**, 1790-1801.

381 Villejoubert, G. & Mandel, D.R. (2002) The inverse fallacy: An account of deviations from Bayes's Theorem
382 and the additivity principle. *Memory & Cognition*, **30**, 171-178.

383 Wintle, B.A., McCarthy, M.A., Parris, K.P. & Burgman, M.A. (2004) Precision and bias of methods for
384 estimating point survey detection probabilities. *Ecological Applications*, **14**, 703-712.

385 Wintle, B.A., Kavanagh, R.P., McCarthy, M.A. & Burgman, M.A. (2005) Estimating and dealing with
386 detectability in occupancy surveys for forest owls and arboreal marsupials. *Journal of Wildlife*
387 *Management*, **69**, 905-917.

388

389

Biosketch

Brendan A. Wintle is a Senior Lecturer in Conservation Science and ARC Future Fellow in the School of Botany at the University of Melbourne, Deputy Director of the National Environment Research Program Environmental Decisions Hub (EDH) and theme leader in the Australian Research Council's Centre of Excellence for Environmental Decisions (CEED). His research focus is in uncertainty and environmental decision making; including evaluating conservation investment effectiveness, optimal monitoring and adaptive management, systematic conservation planning, population viability analysis, species distribution modelling, and biodiversity sampling accounting for imperfect detection. All authors are members of the Quantitative and Applied Ecology group at the University of Melbourne (www.qaeco.com) and participate in EDH (www.aeda.edu.au) and the Australian Research Council's Centre of Excellence for Environmental Decisions. Author contributions: B.W., M.Mc., and T.W. conceived the ideas; K.P. collected the data; M.Mc. and B.W analysed the data; B.W. and M.Mc. led the writing.

Figure legends

Figure 1. A logic tree describing possible outcomes of surveying for a species at 1000 hypothetical locations with imperfect detection. (a) Prior belief of presence; $\Psi' = 0.25$, and probability of detection if the species is present; $p' = 0.5$. (b) The logical structure of the problem when frequencies are converted to probabilities. Note that in our hypothetical example, we assume that the chance of falsely ‘detecting’ an absent species in a single visit (b) is zero. This is a common assumption of most published occupancy and detection models, though this assumption can be relaxed (Bar-Hillel, 1980; Royle & Link, 2006). Inferential outcomes can be classified according to confusion matrix notation (Swets, 1988) as in the last column of the logic tree.

Figure 2. Observation effort required to be 95% sure that a species is absent from a particular site. The Y-axis represents the number of sequential non-detections necessary to be 95% sure the species is absent ($1 - \Psi = 0.05$), the X-axis represents the prior (before data) belief that the species occupies the site (Ψ'), and the three lines correspond to three different prior assumptions about the single-visit detection probability ($p' = 0.1, 0.3, \text{ and } 0.5$), corresponding to the dotted, dashed and solid curves respectively. The prior belief in occupancy could be a subjective probability derived from expert elicitation or a species distribution model fitted to independent data.

Figure 3. Required number of sequential non-detections (Y-axis) to ensure that the probability of *Litoria pearsoniana* absence is > 0.95 as a function of habitat conditions (defined by values of *catchment volume* [X-axis] and the presence or absence of *palms*), and the method of survey (solid line: nocturnal searches, $p = 0.56$; broken line: automatic tape recorders, $p = 0.35$). Plot (a) shows how the required number of surveys varies with catchment volume for sites in which palms are present, and plot (b) gives the same relationship for sites at which palms are absent. The required number of

surveys to be sure of absence is highest for the sites in the best habitat (large streams with palms) when using the least reliable method (automatic tape recorders).

Figure 1.

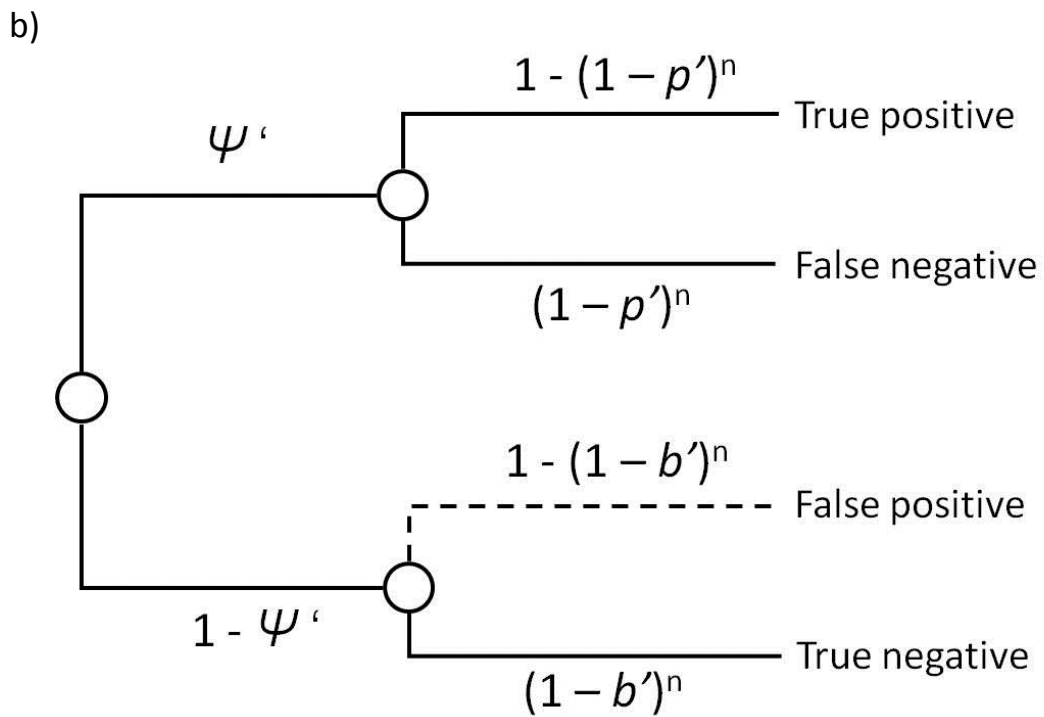
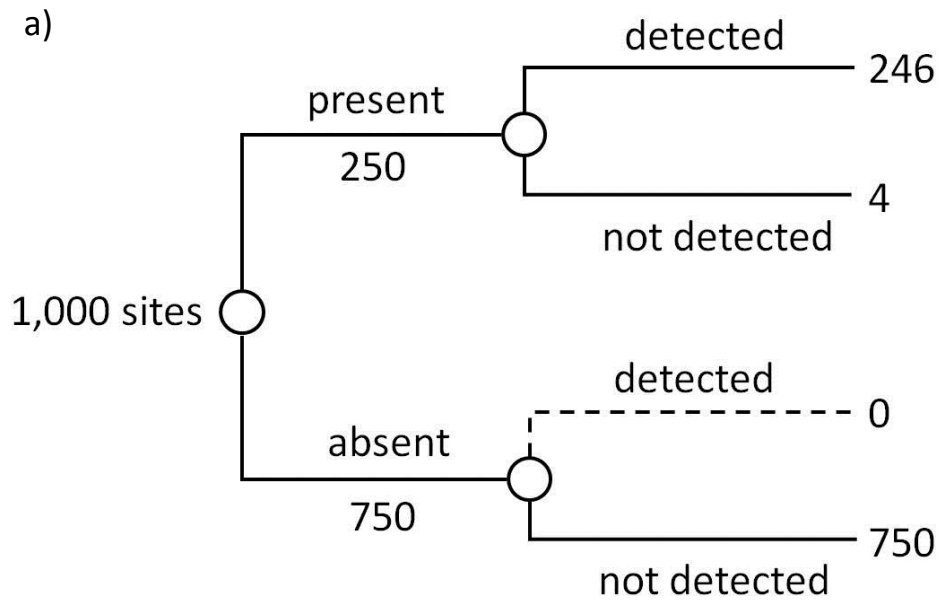


Figure 2

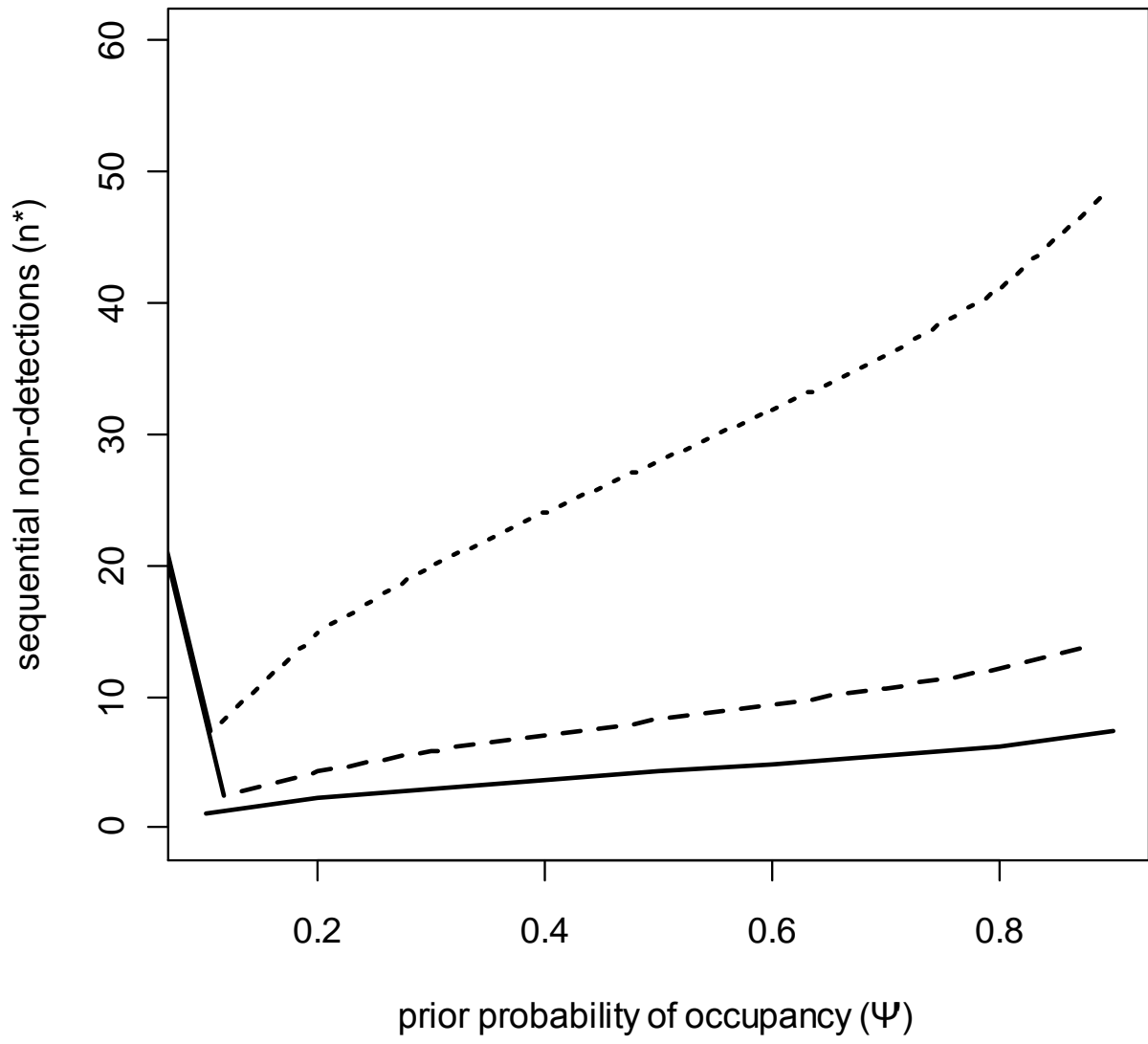
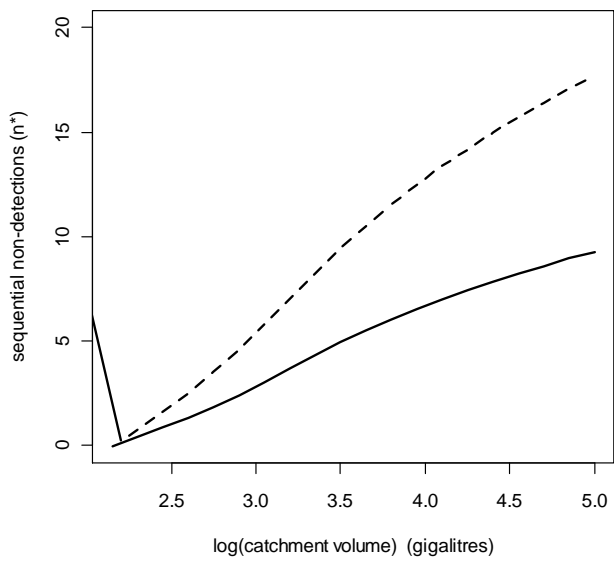


Figure 3.

(a)



(b)

