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

Brendan A. Wintle*[1], Terry V. Walshe[1], Kirsten M. Parris[1] and Michael A. McCarthy[1]

[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

ABSTRACT

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

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

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

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

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

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

(A) Introduction

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

Arguably, the most common application of detectability estimates is in interpreting observation data to determine whether or not a species is, in fact, present at a given site when not detected. Environmental impact assessments utilize these kinds of data to inform decisions about whether or not destruction or development of potential habitats should be allowed to proceed, at the risk of losing endangered species that have not been detected on the site (Garrard *et al.*, 2008). Declaring eradication of a weed or disease depends on the probability that there are unobserved breeding individuals (Regan *et al.*, 2006; Rout *et al.*, 2010). Quarantine operations must assess the probability that an unwanted pest is in fact present in a shipping container, given that it was not detected using a particular search strategy (Burgman *et al.*, 2010). In all such cases, it is important to correctly interpret non-detection data so that decisions can be based on coherent estimates of the probability that a species is truly absent, or alternatively that the species is present but not detected.

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

Bayesian approaches have been applied to modelling imperfect detection data (e.g., Wintle *et al.*, 2005; McCarthy, 2007; Royle *et al.*, 2007; Garrard *et al.*, 2008; Royle & Dorazio, 2008; Burgman *et* *al.*, 2010). The aim of this paper is to demonstrate Bayesian approaches to collecting and interpreting observation data of the kind described above. We provide mathematical expressions and computer code to (i) estimate the probability that a species occupies a given site after one or more non-detections have occurred at that site; (ii) estimate the number of sequential non-detections necessary to assert with a pre-specified degree of confidence that a species is truly absent from the site; and (iii) generalize these to estimate occupancy probabilities at multiple sites, drawing on information about the sorts of places the species has and has not been located over all the sites in a multi-site survey. We demonstrate the application of these methods using a case study based on tree frog survey data from sub-tropical eastern Australia.

(A) Methods

(B) Model

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

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

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

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

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

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

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

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

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

$$n^* => \frac{\log\left(\frac{\Psi}{1-\Psi}\right)-\log\left(\frac{\Psi'}{1-\Psi'}\right)}{\log(1-p')}, \qquad \text{(eq. 3)}$$

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

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

At a site *i* occupied by a species of interest, the likelihood of observing the species in the *j*th visit to the site is p_{ij} and the likelihood of failing to observe the species is $1 - p_{ij}$. Site- and visit-level detection probabilities may vary due to environmental influences on detectability such as vegetation density or visit-level factors such as ambient weather conditions (Wintle *et al.*, 2005), or the survey method used on a given visit to a site (Parris *et al.*, 1999). Let Y_i represent a vector (sequence) of observations of length v_i (the number of visits to the site). Each element of the vector y_{ij} may take the 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 visit. The likelihood of a given detection history (Y_i) over *v* visits to a site *i* is therefore (MacKenzie *et al.*, 2002):

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

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

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

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

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

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

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

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

(B) Application Data

Litoria pearsoniana is a tree frog that breeds in forest streams in sub-tropical eastern Australia. Surveys of 64 sites throughout its range in south-east Queensland and north-east New South Wales were conducted over an area of approximately 14,000 km^2 between 1995-1999 (Parris, 2001). Two survey methods were employed; nocturnal searches and automatic tape recording of advertisement calls. The data comprise a record of the detection or non-detection of the species on each survey night at each survey site. The only visit-level variable considered that could have influenced the probability of detection in each survey is the type of survey method used (search versus tape recording). Variables thought most likely to influence the probability of site occupancy by *L.*

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

(B) Application model

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

(A) Results

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

(B) Minimum survey effort calculations

By utilizing equation three in the OpenBUGS detectability and occupancy model of *L. pearsoniana* (computer code in supplementary material) we were able to estimate the minimum survey effort (number of repeat visits) necessary to achieve some pre-specified confidence in a conclusion that the species was truly absent from a particular location under a range of environmental and detectability conditions that may be encountered in future surveys (Fig. 3). It is apparent from Figure 3 that the number of sequential non-detections necessary to be 95% sure that the species is absent from a given site increases as the variables that positively influence probability of *L. pearsoniana* occupancy increase. Under the most effective survey method (spotlighting streams), a 10-fold increase in the volume of streams with palms leads to a 2-3-fold increase in the number of non-detections necessary to be 95% certain that the species is, in fact, absent. This is because the prior probability that the species occupies larger streams is substantially higher than that for smaller streams, necessitating a greater weight of evidence (in the form of sequential non-detections) to provide the same level of (posterior; after data) confidence that the species is absent. At the highest level of catchment volume recorded in the study, in a site containing palms, approximately 9 sequential non-detections using spotlighting surveys are required to be 95% sure the site is unoccupied, compared with the 18 non-detections using tape recording that would be needed for the same level of confidence in absence (Fig 3).

(A) Discussion

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

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

The insights from models developed here emphasise the importance of *explicit* estimation and use of prior beliefs. Estimates may be based directly on biological judgment, the predictions generated from a habitat model, or simply the *unconditional rate of occupancy* (also known as expected prevalence or the 'base rate') from previous surveys of the species (MacKenzie, 2005). An uninformative prior probability of occupancy, Ψ' = 0.5 may be difficult to justify in many instances. For example, consider a species that on the basis of historical records is estimated to be present at 10% of sites within a study region of variable habitat quality. If p = 0.3 and we wish to be 99% confident of absence, then from equation 3, we require 13 sequential non-detections should we insist on use of the uninformative prior, Ψ' = 0.5. If we use Ψ' = 0.10, then seven non-detections are required. When surveying resources are scarce, use of an uninformative prior represents an opportunity cost. A sophisticated approach that recognises opportunity costs would utilise biological judgment to discern areas where the species is more (or less) likely to be present than the overall 10% estimate of prevalence.

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

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

(A) Acknowledgements

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

(A) References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 at1000 hypothetical locations with imperfect detection. (a) Prior belief of presence; Ψ' = 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-Ψ = 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, 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.

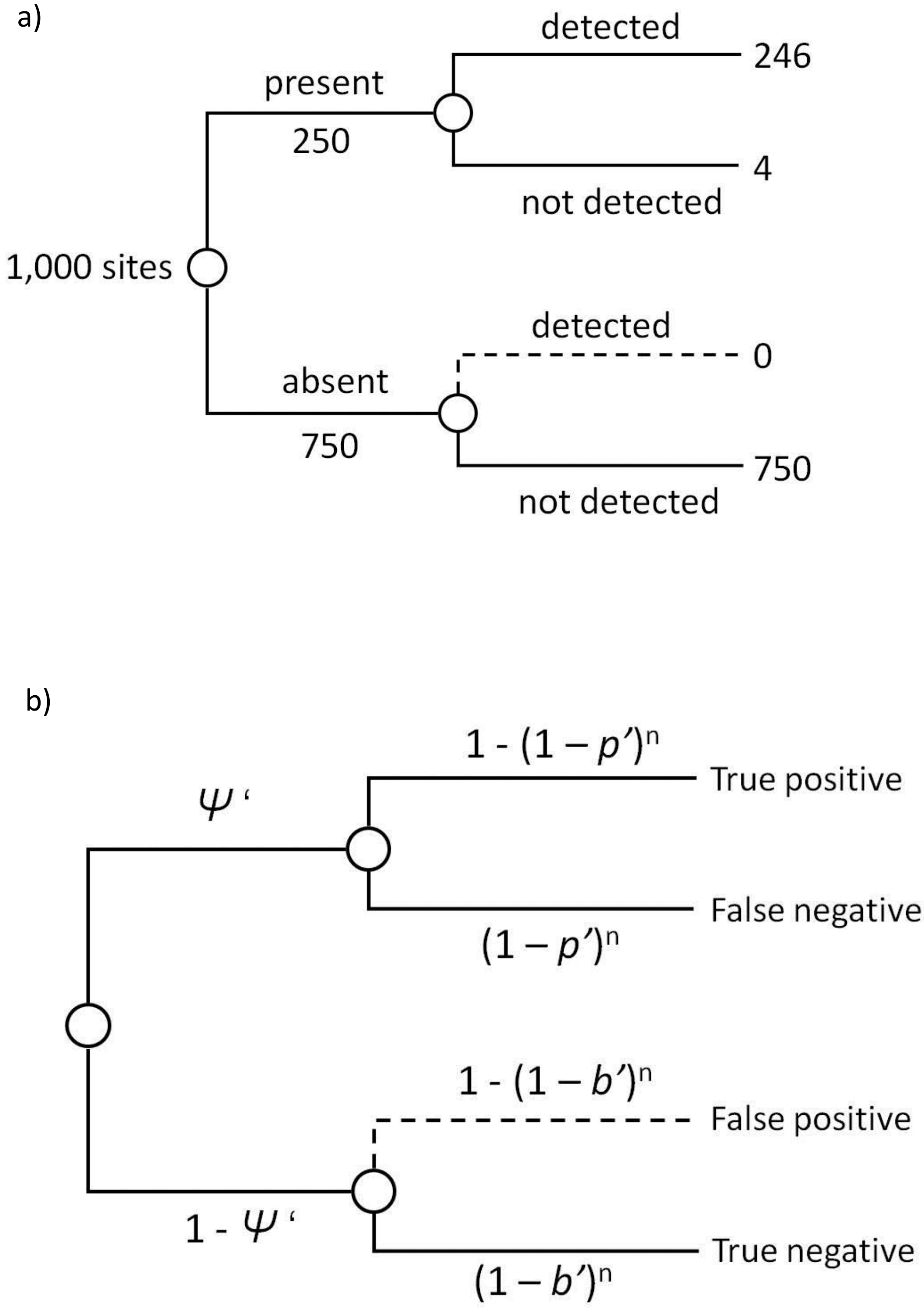
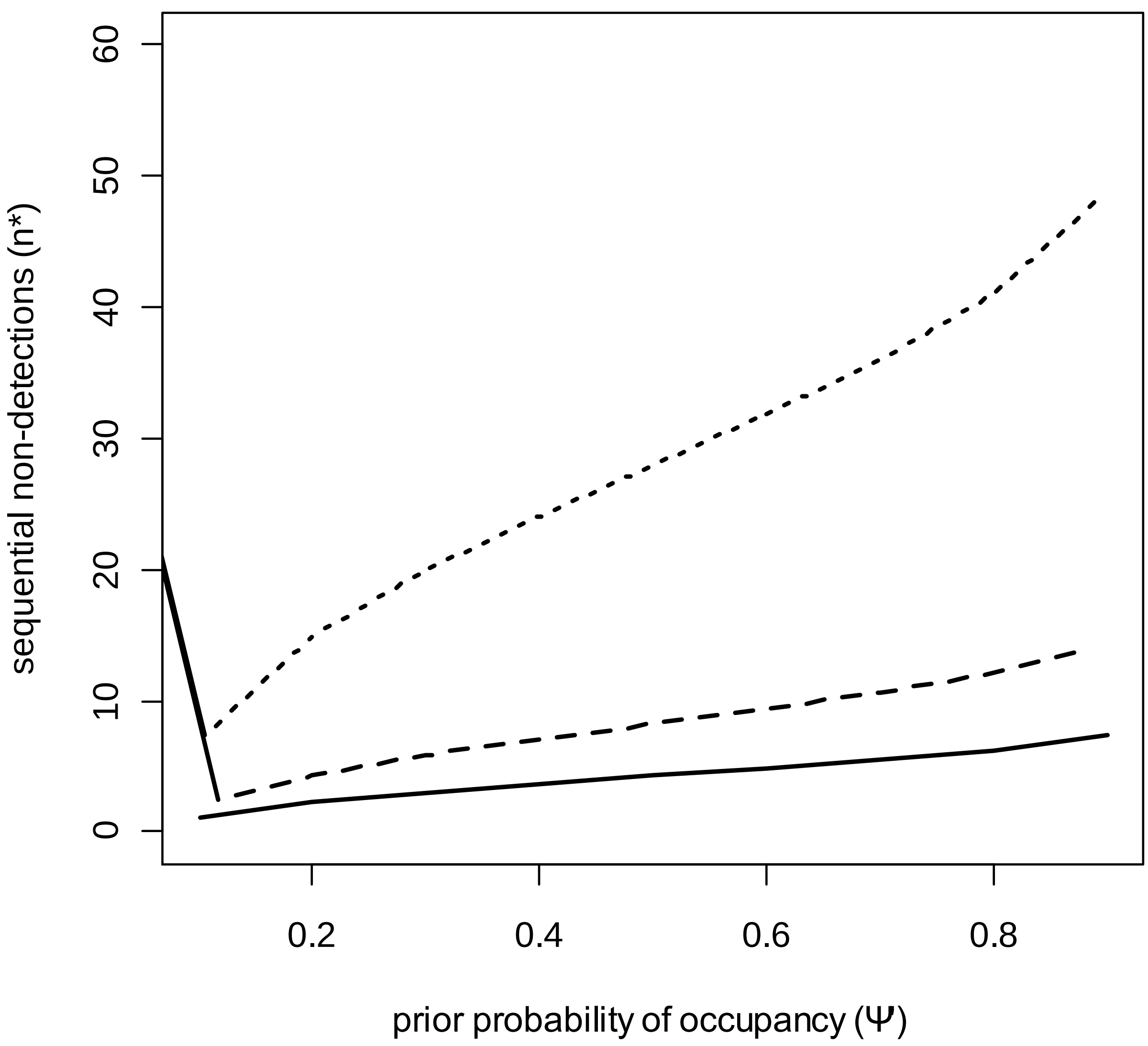

a)
detected
246
present
250
4
not detected
1,000 sites
detected
0
absent
750
750
not detected
b)
$1 - (1 - p')^n$
True positive
Ψ'
False negative
$(1 - p')^n$
$1 - (1 - b')^n$
False positive
$1 - \Psi'$
True negative
$(1 - b')^n$

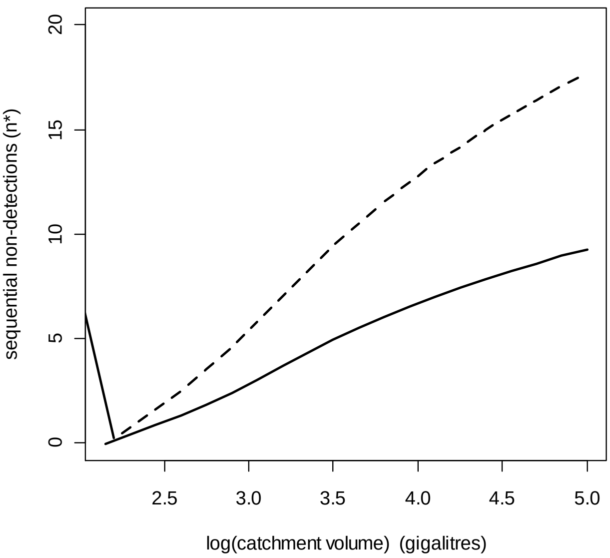
Figure 2



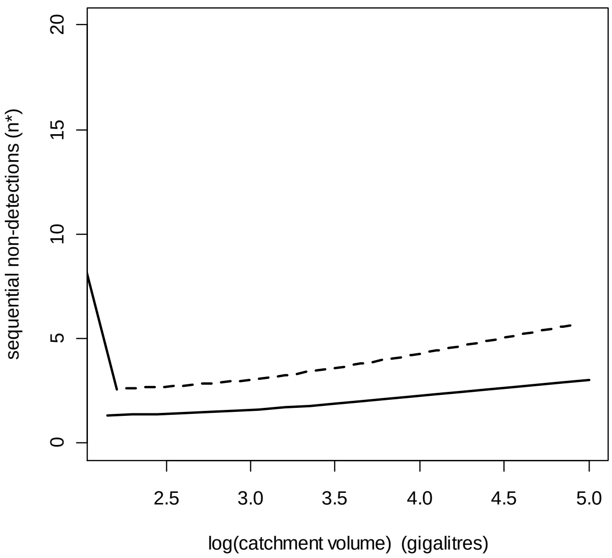
60
50
40
30
20
10
0
sequential non-detections (n*)
0.2
0.4
0.6
0.8
prior probability of occupancy (Ψ)

Figure 3.

(a)


sequential non-detections (n*)
0
5
10
15
20
2.5
3.0
3.5
4.0
4.5
5.0
log(catchment volume) (gigalitres)

(b)


sequential non-detections (n*)
0
5
10
15
20
2.5
3.0
3.5
4.0
4.5
5.0
log(catchment volume) (gigalitres)