Occupancy estimation and the closure assumption

Christopher T. Rota*,[1], Robert J. Fletcher Jr[1], Robert M. Dorazio[2] and Matthew G. Betts[3]

[1]*Wildlife Ecology and Conservation, University of Florida, PO Box 110430, Gainesville, FL 32611-0430, USA;*
[2]*Department of Statistics, US Geological Survey and University of Florida, Gainesville, FL 32611-0339, USA; and*
[3]*Department of Forest Ecosystems and Society, Oregon State University, 216 Richardson Hall, Corvallis, OR 97331, USA*

Summary

1. Recent advances in occupancy estimation that adjust for imperfect detection have provided substantial improvements over traditional approaches and are receiving considerable use in applied ecology. To estimate and adjust for detectability, occupancy modelling requires multiple surveys at a site and requires the assumption of 'closure' between surveys, i.e. no changes in occupancy between surveys. Violations of this assumption could bias parameter estimates; however, little work has assessed model sensitivity to violations of this assumption or how commonly such violations occur in nature.

2. We apply a modelling procedure that can test for closure to two avian point-count data sets in Montana and New Hampshire, USA, that exemplify time-scales at which closure is often assumed. These data sets illustrate different sampling designs that allow testing for closure but are currently rarely employed in field investigations. Using a simulation study, we then evaluate the sensitivity of parameter estimates to changes in site occupancy and evaluate a power analysis developed for sampling designs that is aimed at limiting the likelihood of closure.

3. Application of our approach to point-count data indicates that habitats may frequently be open to changes in site occupancy at time-scales typical of many occupancy investigations, with 71% and 100% of species investigated in Montana and New Hampshire respectively, showing violation of closure across time periods of 3 weeks and 8 days respectively.

4. Simulations suggest that models assuming closure are sensitive to changes in occupancy. Power analyses further suggest that the modelling procedure we apply can effectively test for closure.

5. *Synthesis and applications.* Our demonstration that sites may be open to changes in site occupancy over time-scales typical of many occupancy investigations, combined with the sensitivity of models to violations of the closure assumption, highlights the importance of properly addressing the closure assumption in both sampling designs and analysis. Furthermore, inappropriately applying closed models could have negative consequences when monitoring rare or declining species for conservation and management decisions, because violations of closure typically lead to overestimates of the probability of occurrence.

Key-words: breeding birds, colonization, detection probability, dynamic models, extinction, metapopulation, point counts, removal models, robust design

Introduction

Estimating and interpreting patterns of occupancy lie at the heart of many questions in ecology and problems in conservation. For example, metapopulation theory often explores variation in patch occupancy in fragmented landscapes (Hanski 1994). Species distribution models, which are widely used in guiding conservation and management decisions, frequently rely on observed patterns of detection and non-detection (Guisan *et al.* 2006). Additionally, occupancy can provide valuable information on population trends when more detailed demographic or abundance estimates are not practical (Bailey, Simons & Pollock 2004).

Traditional approaches to occurrence estimation, such as logistic regression or Incidence Function Models (Hanski 1994), assume perfect detection of species. Recently, these approaches have been criticized because even modest amounts of false absences (i.e. modelling a species as absent

*Correspondence author. E-mail: ctr4g2@mail.missouri.edu

when it is in fact present) can bias parameter estimates in metapopulation models and predicted habitat relationships (Moilanen 2002; Tyre *et al.* 2003; Gu & Swihart 2004; Martin *et al.* 2005).

Because of the bias introduced by non-detection errors, several recent investigations have focused on how to model occupancy, given imperfect detection (Geissler & Fuller 1987; Azuma, Baldwin & Noon 1990; MacKenzie *et al.* 2002; Tyre *et al.* 2003; Stauffer, Ralph & Miller 2004). Of these new techniques, MacKenzie & Royle (2005) suggest that the approach of MacKenzie *et al.* (2002) is the most flexible and that other approaches are special cases of their general model. This approach has provided significant improvements over traditional approaches, which is reflected in a recent surge in occupancy studies (Marsh & Trenham 2008).

To estimate detection probability, MacKenzie *et al.*'s (2002, 2006) occupancy-modelling approach requires multiple surveys at each site. Detection probability is then estimated from the pattern of detections and non-detections that arise from these multiple surveys. A necessary assumption for estimating and accounting for detectability is that sites are closed to changes in occupancy between surveys, which has been described as the 'closure assumption'. The term 'closure' reflects the assumption that if a site is occupied during at least one survey, it is assumed to have been occupied during all surveys, and any non-detection during a survey is considered a 'false zero' or a 'false negative'.

A commonly used sampling approach for occupancy studies is to visit a site multiple times and conduct a single survey during each site visit (e.g. Bailey, Simons & Pollock 2004; Ball, Doherty & McDonald 2005). We will refer to this approach as a *standard occupancy* sampling protocol. Site visits are frequently separated by periods of weeks or months, and sites are assumed to be closed during these time periods. Violations of this assumption may lead to biased estimates of occupancy. However, little work has been carried out to assess how violations of the closure assumption may affect occupancy estimates, and MacKenzie *et al.* (2006) have only inferred the strength and direction of these biases from Kendall's (1999) evaluation of capture–recapture models. Although MacKenzie *et al.* (2006) provide useful suggestions for reducing the problem of closure, no formal framework has been developed to explicitly test the closure assumption.

Here, we address the closure assumption for occupancy estimation by advocating the use of Pollock's (1982) robust design over short time intervals, wherein an observer conducts multiple surveys during each site visit. We show that this approach permits estimation of transitions in site occupancy (i.e. local colonization and extinction) and formal statistical tests of closure between site visits. Using two data sets on bird distributions, we test the likelihood of closure over time-scales typical of many wildlife occupancy investigations. Finally, using simulations, we assess how sensitive occupancy models are to violations of the closure assumption and evaluate the power of likelihood-ratio tests to identify these violations.

Materials and methods

SAMPLING AND MODELS FOR ESTIMATING CLOSURE

An intuitive and practical way to address the closure assumption is to sample populations using the robust design, which was originally developed for capture–recapture sampling. In the robust design, sampling consists of secondary sampling periods nested within primary sampling periods. Populations are assumed to be closed to demographic changes between secondary sampling periods and open to demographic changes between primary sampling periods. We apply principles of robust-design (RD) sampling in an occupancy context by considering individual site-visits as primary sampling periods and multiple surveys conducted during each site-visit as secondary sampling periods. We define 'site-visit' as any single visit to a site during which one or more surveys are conducted to assess detection/non-detection. Thus, we assume that sites are open to changes in occupancy between site-visits, but are closed to changes in occupancy during site-visits. Conducting multiple surveys during each site-visit minimizes the time over which closure is assumed while still providing the detection and non-detection data necessary to estimate detection probability (cf. MacKenzie *et al.* 2006).

Here, we define two RD sampling protocols. A *fixed-replicate* RD protocol consists of conducting a fixed number of independent surveys during each primary sampling period, regardless of detection history. With a fixed-replicate RD protocol, single site-visit estimates of occupancy can be computed using the approach of MacKenzie *et al.* (2002). When conducting multiple surveys in rapid succession, assuring independence between surveys may prove problematic. For example, a species detected during one survey may be easier to detect during subsequent surveys once its location is known. For that reason, MacKenzie *et al.* (2006) suggest adopting a 'removal' sampling protocol, which consists of surveying for a species only until it is first detected, up to a maximum of *J* surveys (Azuma, Baldwin & Noon 1990; MacKenzie & Royle 2005). As surveying stops once a species is first detected, assuming independence between surveys is less problematic. We refer to this approach as a *removal* RD protocol. With a removal RD protocol, single site-visit estimates of occupancy can be computed using MacKenzie *et al.*'s (2006, p. 102) single-season removal model.

Transitions in occupancy between primary sampling periods can be estimated by fitting dynamic models to data collected using RD protocols. These models include the probability of local colonization (γ) and extinction (ε) as parameters; γ is the probability that an unoccupied site at time t will become occupied at time $t + 1$, and ε is the probability that an occupied site at time t will become unoccupied at time $t + 1$. Data collected with a fixed-replicate RD protocol can be fit to MacKenzie *et al.*'s (2003) dynamic occupancy models. Data collected with a removal RD protocol can be fit using a simple extension of MacKenzie *et al.*'s (2006, p. 102) single-season removal model. The likelihood function for a two-season, dynamic removal model is:

$$L(\psi_{int}, p_1, p_2, \varepsilon, \gamma | j_{i1}, j_{i2}, y_{i1}, y_{i2})$$

$$= \prod_{i=1}^{N} \{\psi_{int} p_1^{y_{i1}} (1 - p_1)^{j_{i1} - y_{i1}} [(1 - \varepsilon) p_2^{y_{i2}} (1 - p_2)^{j_{i2} - y_{i2}} + \varepsilon I(y_{i2} = 0)]$$

$$+ (1 - \psi_{int}) I(y_{i1} = 0) [\gamma p_2^{y_{i2}} (1 - p_2)^{j_{i2} - y_{i2}} + (1 - \gamma) I(y_{i2} = 0)]\},$$

$$\text{eqn 1}$$

where ψ_{int} is the probability a site is occupied during time 1; p_1 and p_2 are the conditional probabilities of detecting a species, given presence, during times 1 and 2 respectively; j_{i1} and j_{i2} denote the number of surveys until the first detection of a species at site i during times 1 and 2 respectively (note that if a species

remains undetected at site i during time t, then $j_{it} = J$); y_{i1} and y_{i2} are binary indicators of whether a species is detected ($y = 1$) or not ($y = 0$) at site i during times 1 and 2 respectively; I is an indicator function ($I = 1$ if y_{i1} or $y_{i2} = 0$, $I = 0$ otherwise); and N is the number of sites surveyed. The framework for extending this dynamic removal model to three or more seasons is identical to MacKenzie *et al.*'s (2003) dynamic model, changing only the observation component of the model to reflect a removal sampling protocol.

A likelihood-ratio comparison of open (ε and/or $\gamma > 0$) and closed ($\varepsilon = \gamma = 0$) models can be used to formally test for closure between primary sampling periods. This test is a ratio of the likelihoods of two nested models (Λ) calculated using the maximum-likelihood estimate (MLE) of parameters under the null hypothesis ($\varepsilon = \gamma = 0$) and the MLE of parameters under the alternative hypothesis (ε, $\gamma > 0$). The likelihood-ratio test statistic is calculated as $-2 \times \log(\Lambda)$. Under the standard regularity conditions, the limiting distribution the test statistic under the null hypothesis is χ^2 with degrees of freedom equal to the difference in dimensionality (parameters) between the two models (Royle & Dorazio 2008, p. 65). However, in this situation, the null model has parameters that are on the boundary of the parameter space ($\varepsilon = \gamma = 0$), and the limiting distribution is a mixture of χ^2 and zeros (Self & Liang 1987). The mixing proportion of this distribution depends on the Fisher information matrix and is difficult to calculate; however, the distribution of the test statistic can be approximated by simulating hypothetical data sets under the null hypothesis (Appendix S1).

APPLICATION TO BREEDING BIRD DISTRIBUTIONS

As an application, we fit standard occupancy and RD models to avian point-count data. We used two separate data sets that illustrate different RD sampling designs. Our first data set was collected along the Madison and Missouri Rivers, Montana, USA, during the 2004 breeding season (Fletcher & Hutto 2008; 'riparian' data set hereafter). Each of 165 sites was visited twice, once between 25 May and 15 June and again between 15 June and 10 July, with an average of 3 weeks between visits. A standard 10-min, 50-m radius point-count survey was conducted during each site visit. Each 10-min survey was further divided into four 2·5-min sampling intervals, which served as secondary sampling periods. This data set was collected using a removal RD sampling protocol, so sampling stopped for a species during a site-visit once it was detected.

Our second data set was collected at the Hubbard Brook Experimental Forest, New Hampshire, USA, during the 2007 breeding season (Betts *et al.* 2008; 'Hubbard Brook' data set hereafter). Each of 184 sites was visited three times between 2 June and 2 July, with 6–8 days between visits. A standard 10-min, 50-m radius point-count was conducted during each site visit. Each 10-min survey was further divided into three 3-min 20-s sampling intervals, which served as secondary sampling periods. This data set was collected using a fixed-replicate RD sampling protocol, so each species was re-sampled during each sampling interval.

We fit open and closed RD models to both data sets. We treated closed RD models as a restricted version of open RD models, such that $\varepsilon = \gamma = 0$. We also fit standard occupancy models to a truncated version of each data set. We generated truncated data sets by collapsing information from each primary sampling period into either a 1 if a species was detected or a 0 if a species remained undetected. This effectively treated each site visit as a single survey.

Several considerations should be made when determining how to model detection probability (p) within and between site visits. For closed models, allowing estimates of p to vary between site visits has the potential to 'absorb' violations of the closure assumption (MacKenzie *et al.* 2006). Open models allow transitions in occupancy between site-visits and thus do not need to 'absorb' violations of closure. Nonetheless, allowing p to vary between site visits in open models may provide a means to distinguish changes in site occupancy vs. changes in detectability. Differences in p could also be modelled within site visits.

We fit models that assumed both constant p and which allowed p to vary between site visits. We assumed constant p within site-visits because each site-visit was only 10 min in duration. This resulted in fitting four RD models for each species: closed with constant p, closed with site-visit-specific p, open with constant p, and open with site-visit-specific p. For both data sets, we fit models to species that were detected on > 10% of sites surveyed (riparian = 28 species, Hubbard Brook = 18 species).

We used both likelihood-ratio comparisons and Bayesian information criterion (BIC) to evaluate closure for both data sets. Likelihood-ratio comparisons provide formal tests for closure and facilitate the use of power analyses (see below), but comparisons can only be made between nested models. BIC complements the likelihood-ratio comparison because it allows comparison of non-nested models, but provides no formal test for closure. Additionally, BIC provides a conservative measure of support for open models relative to the more frequently used Akaike's information criterion, which is known to favour highly parameterized models in many cases (Link & Barker 2006). We tested for closure using likelihood-ratio comparisons of nested open and closed RD models. We calculated the BIC for each RD model as $-2\log(L) + K \times \log(N)$, where K denotes the number of estimable parameters (Link & Barker 2006). We then computed a set of BIC model weights for each model's relative fit for each species following Link & Barker (2006).

We define three different measures of the probability of occurrence to facilitate comparison between open and closed models. First, we define ψ as the probability of occurrence from closed models. Second, we define ψ_{int} as the probability of occurrence during the first primary sampling period for open RD models. Finally, we define $\psi_{breeding}$ as the probability that a site is occupied at least once during a time period of interest (e.g. the breeding season), which may encompass more than one primary sampling period and accounts for the potential for populations to be open during this time period. We make this distinction because ψ, estimated from closed models, and ψ_{int}, estimated from open models, are on different time-scales. For example, estimates of ψ_{int} from open models only applies to the first primary sampling period, while estimates of ψ from closed models apply to all primary sampling periods. To compare open and closed estimates of the probability of occurrence over similar time-scales, we thus calculate $\psi_{breeding}$ using parameters estimated from open models. We calculated $\psi_{breeding}$ for the riparian data set as:

$$\psi_{breeding} = \psi_{int} + (1 - \psi_{int}) \times \gamma.$$

Similarly, we calculated $\psi_{breeding}$ for the Hubbard Brook data set as:

$$\psi_{breeding} = \theta + (1 - \theta) \times \gamma_2,$$

where $\theta = \psi_{int} + (1 - \psi_{int}) \times \gamma_1$. This approach can be extended via recursion for any number of primary time periods. We used a parametric bootstrap procedure to estimate 95% confidence intervals for $\psi_{breeding}$.

To assess how violations of the closure assumption can more generally bias parameter estimates, and to interpret the statistical power of detecting violations of closure, we conducted simulation studies with sites open and closed to changes in occupancy between primary sampling periods. We simulated high and low initial probability of site occupancy ($\psi_{int} = \{0.3, 0.7\}$) and single-site-visit detection probability ($p = \{0.3, 0.7\}$), for a total of four combinations of occupancy and detection probability. We chose to not focus on lower values of p because MacKenzie *et al.* (2002) found that standard occupancy models were biased when $p < 0.3$; however, we further explored more extreme variation in p for a subset of parameter combinations (see below). For each combination of ψ_{int} and p, we simultaneously varied both colonization and extinction between primary sampling periods ($\gamma, \varepsilon = \{0.00-0.95\}$).

To guide our simulations, we adopted the removal and fixed-replicate sampling protocols used by Fletcher & Hutto (2008) and Betts *et al.* (2008) respectively. We based one set of simulations on Fletcher & Hutto (2008) by simulating observations from two primary sampling periods, each divided into four secondary sampling periods ($J = 4$ surveys). We based a second set of simulations on Betts *et al.* (2008) by simulating observations from three primary sampling periods, each divided into three secondary sampling periods ($J = 3$ surveys).

For each replicate data set, we simulated observations on $N = 1000$ and $N = 150$ sites. Each simulated site was initially occupied with probability ψ_{int}, and species were detected on occupied sites with probability $1 - (1 - p)^{1/J}$ during each secondary sampling period. During subsequent primary periods, species were absent from sites occupied during the previous primary period with probability $1 - (1 - \varepsilon)^{1/(s-1)}$, where s is the number of primary sampling periods, and present on sites that were unoccupied during the previous primary period with probability $1 - (1 - \gamma)^{1/(s-1)}$ (note that total colonization and extinction rates were equal for scenarios with two or three primary sampling periods). We randomly generated 1000 replicate data sets for each combination of ψ_{int}, p, ε, γ and N. For a subset of parameter combinations (removal sampling design and $N = 1000$), we also simulated declining detection probabilities between primary sampling periods (p declining from 0.7 to 0.3 or to 0.1) to interpret whether open models could distinguish violation of closure from changes in detection probability.

Using the data-generating process described above, we simulated three different sampling protocols to address the closure assumption. For a fixed-replicate RD protocol, each primary sampling period consisted of J secondary sampling periods. For a removal RD protocol, we only surveyed during a primary sampling period until a species was first detected, for a maximum of J surveys. For a standard occupancy protocol, we truncated information from primary sampling periods into a 0 for non-detection and 1 for detection, effectively treating each primary sampling period as a single survey.

We then fit open and closed RD models for both sampling protocols. Fixed-replicate RD models had $J \times s$ surveys. Removal RD models had a maximum of $J \times s$ surveys, but surveys stopped after the initial detection in any primary sampling period. For the standard occupancy sampling protocol, we fit single-season occupancy models (MacKenzie *et al.* 2002) to the truncated data, with the number of surveys equal to the number of primary sampling periods. We allowed detection probability to vary between primary sampling periods for all models, which MacKenzie *et al.*

(2006) suggested may 'absorb' some violations of the closure assumption.

We concluded by calculating per cent relative bias of closed models for each simulated scenario. We calculated per cent relative bias as bias $= (E(\psi) - \psi_{breeding})/\psi_{breeding}$, where $E(\psi)$ is the average estimated ψ for closed models from all simulations, and $\psi_{breeding}$ is calculated as above using the 'true' values of ψ_{int} and γ. Consequently, $\psi_{breeding}$ is the true value of occupancy across the time period of interest.

Results

### BREEDING BIRD DISTRIBUTIONS

For 16 of the 28 species considered from the riparian data set, open RD models received more than half of the model weight, according to BIC weights (Fig. 1a). Likelihood-ratio tests provided similar results, rejecting the null hypothesis of closure between primary sampling periods for 20 species ($P < 0.05$). In general, apparent colonization or extinction events were best explained by transitions in occupancy rather than changes in detectability; i.e. seasonal changes in detection probability were insufficient for explaining apparent transitions. In only two clear cases (house finch *Carpodacus mexicanus* and European starling *Sturnus vulgaris*) did closed models with site-visit-specific detection probability best explain apparent changes in site occupancy (Fig. 1a). In all cases, estimates of $\psi_{breeding}$ calculated from open RD models were lower than estimates of ψ from standard occupancy models (Fig. 2a; mean difference = 0.15).

The null hypothesis of closure tended to be supported in situations where estimated detection probability from open models was low. For example, black-headed grosbeak *Pheucticus melanocephalus*, downy woodpecker *Picoides pubescens* and American goldfinch *Carduelis tristis*, which all received substantial support from closed models (Fig. 1), also had the lowest estimated detection probabilities from open models. This highlights the difficulty associated with distinguishing between non-detection and changes in site occupancy when p is low. This result also demonstrates that tests for closure are conservative when p is low. Conversely, species with low estimates of ε or γ (see Appendix S2 for estimates of γ and ε) for which the null hypothesis of closure was rejected, such as common yellowthroat *Geothlylpis trichas* or house wren *Troglodytes aedon*, tended to have high estimated detection probabilities from open models.

For 17 of 18 species considered from the Hubbard Brook data set, open RD models received more than half of the model weight, according to BIC weights (Fig. 1b). Likelihood-ratio tests provided similar results, rejecting the null hypothesis of closure between site visits for all 18 species ($P < 0.01$). Apparent colonization and extinction events with the Hubbard Brook data set were always better explained by transitions in occupancy rather than changes in detection probability (Fig. 1b). In all cases, estimates of $\psi_{breeding}$ calculated from open RD models were lower than estimates of ψ from standard occupancy models (Fig 2b; mean difference = 0.37).

(a)

(b)
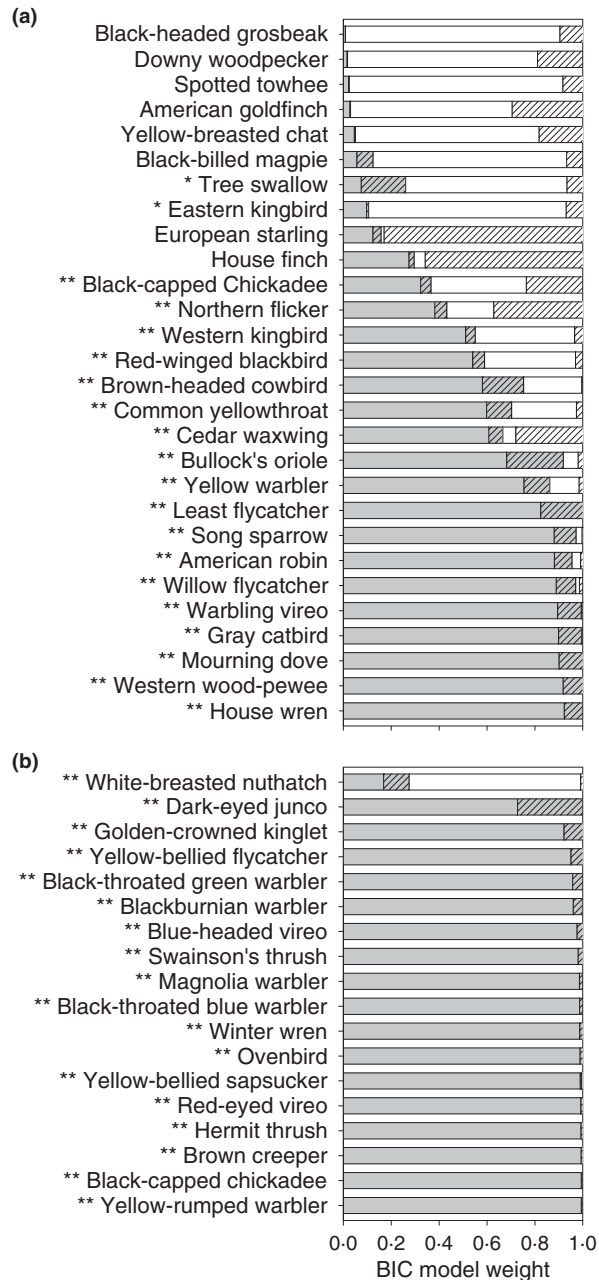
Fig. 1. Relative weights of open and closed occupancy models for breeding birds in two areas. (a) Bayesian information criteria (BIC) model weight for all species detected on > 10% of riparian point-counts (N = 165), Montana, USA, using a removal RD sampling protocol. (b) BIC model weight for all species detected on > 10% of Hubbard Brook point-counts (N = 184), New Hampshire, USA, using a fixed-replicate sampling protocol. Grey bars indicate BIC weights for open models and white bars indicate BIC weights for closed models. Hatched bars indicate BIC weights for models with site-visit-specific detection probability and bars without hatching indicate BIC weights for models with constant detection probability. Significant likelihood-ratio tests for closure are marked with asterisks (*P < 0.05, **P < 0.01). We report likelihood-ratio tests for nested RD models calculated with either constant detection probability or site-visit-specific detection probability, whichever has the greatest support according to BIC weights. For a list of scientific names, see Appendix S2.

(a)

(b)
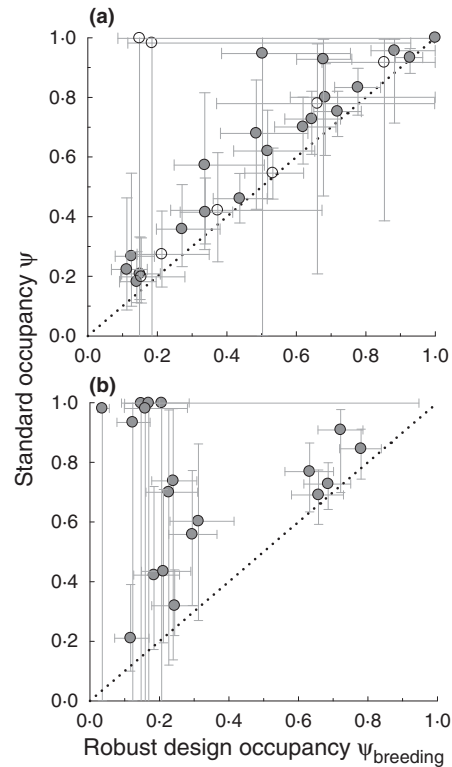
Fig. 2. (a) Estimated probability of site occupancy, $\psi_{breeding}$ (±95% CI), from open removal RD models relative to standard occupancy models for all species detected on > 10% of riparian data set point-counts. Open circles indicate species for which the closure hypothesis was supported. (b) Estimated $\psi_{breeding}$ (±95% CI), from open fixed-replicate RD models relative to standard occupancy models for all species detected on > 10% of Hubbard Brook data set point-counts. The dotted line indicates no difference in estimates, and points above that line indicate a higher estimate of ψ for standard occupancy models. We report point-estimates from models calculated with either constant detection probability or site-visit-specific detection probability, whichever had the greatest support according to Bayesian information criteria weights.

SIMULATION STUDY

Closed model estimates of ψ are sensitive to changes in occupancy between primary sampling periods. For example, average bias of ψ with two primary periods and parameter combinations shown in Fig. 3 is 0.19 (SD = 0.25). With two primary sampling periods, closed model estimates of ψ were unbiased if extinction or colonization-only events occurred. With three primary sampling periods, closed model estimates of ψ were slightly negatively biased if extinction or colonization-only events occurred.

Open models generally provided unbiased estimates of ψ_{int} and $\psi_{breeding}$. For example, average bias of $\psi_{breeding}$ with two primary periods and parameter combinations shown in Fig. 3 was 0.004 (SD = 0.009). Additionally, open models were generally able to distinguish between changes in occupancy and changes in detection probability between primary sampling periods: when $\gamma = \varepsilon = 0$, but p declined from 0.7 to 0.3, average bias in estimates of $\psi_{breeding}$ was 0.01 for open models (0.003 for closed
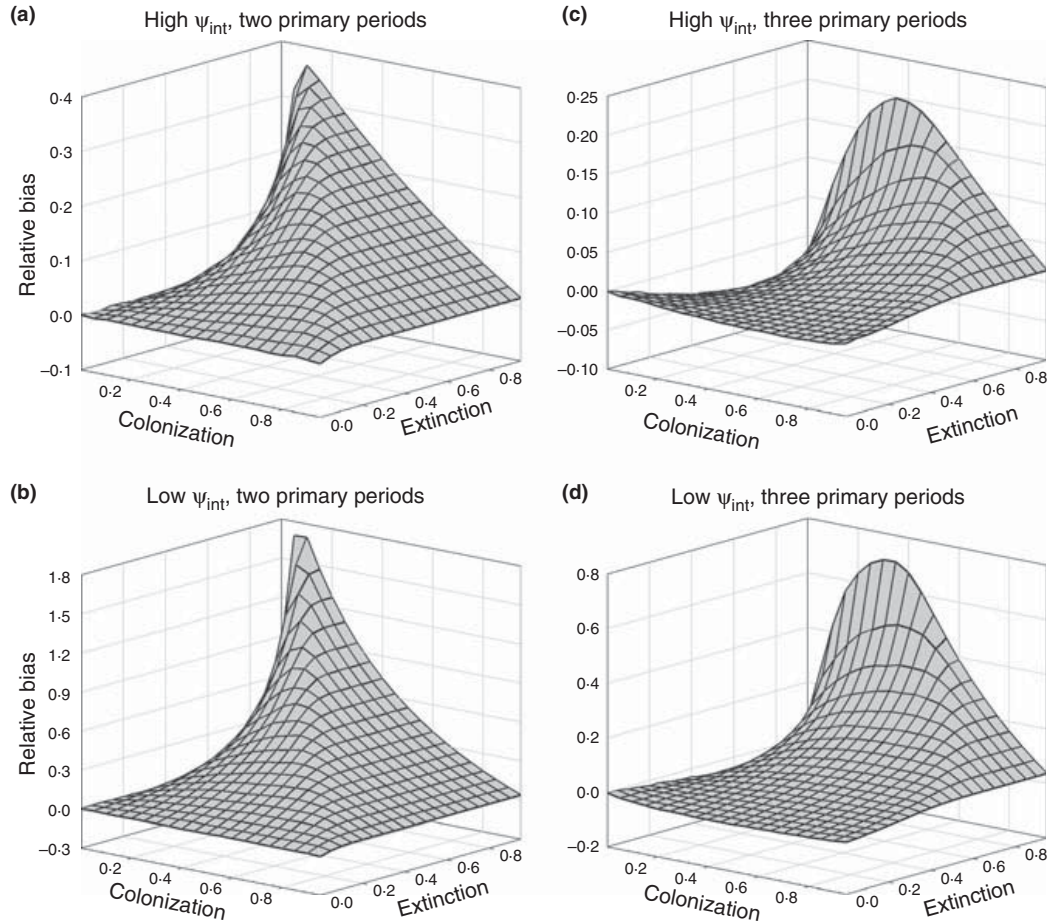
Fig. 3. Estimated relative bias from standard occupancy models (MacKenzie *et al.* 2002), allowing estimates of detection probability, *p*, to vary between primary sampling periods. Mean estimates were calculated from 1000 replicate simulations. High $\psi_{int} = 0.70$ and low $\psi_{int} = 0.30$. For all simulations, $p = 0.7$ and $N = 1000$. Note that bias from open models is not shown, because bias is approximately zero (see text).

models). Open models tend to overestimate $\psi_{breeding}$ when both ψ_{int} is low and ε is high. This bias is minimal at high values of *p* and *N*, and increases as both *p* and *N* decrease.

When colonization and extinction vary simultaneously, closed models are usually biased high. This bias is most pronounced at intermediate values of γ, and increases with increasing ε (Fig. 3). Additionally, closed model bias is negatively correlated with ψ_{int}. Estimates of ψ from closed models were less biased with three primary sampling periods than with two primary sampling periods (Fig. 3).

Although closed model estimates of ψ varied depending on the exact combination of *N*, *p*, ψ_{int}, γ, ε, or sampling protocol, the same general pattern described above held for all scenarios (C.T. Rota, unpublished results). Closed model estimates of ψ were, on average, similar with $N = 1000$ and $N = 150$. Additionally, closed model estimates of ψ were, on average, similar with $p = 0.7$ and $p = 0.3$. Exploration of more extreme detection probabilities ($p = 0.1$ and $p = 0.9$) on a subset of parameter combinations revealed a similar pattern of biases. Both low sample size and low detection probability increased the uncertainty of the estimate, primarily leading to a response surface

that was less smooth than shown in Fig. 3. Estimates of ψ from closed standard occupancy models were more biased than closed RD models, with removal models demonstrating more bias than fixed-replicate models (C.T. Rota, unpublished results).

Importantly, the power of a likelihood-ratio test to detect violations of the closure assumption increases with increasing extinction and colonization (Fig. 4, see Appendix S1 for power calculations). Several factors affect the power to detect a violation of closure. Power is slightly greater for a fixed-replicate RD sampling protocol than for a removal RD sampling protocol (not shown), and power is greater with three sampling periods than with two sampling periods (Fig. 4). Interestingly, power decreases at very high levels of extinction (Fig. 4), presumably because of an inability to distinguish between extinction and a decrease in detection probability.

Discussion

Occupancy estimation is valuable for many applications in ecology and conservation. Yet our results suggest that application of sampling designs and models aimed at estimating
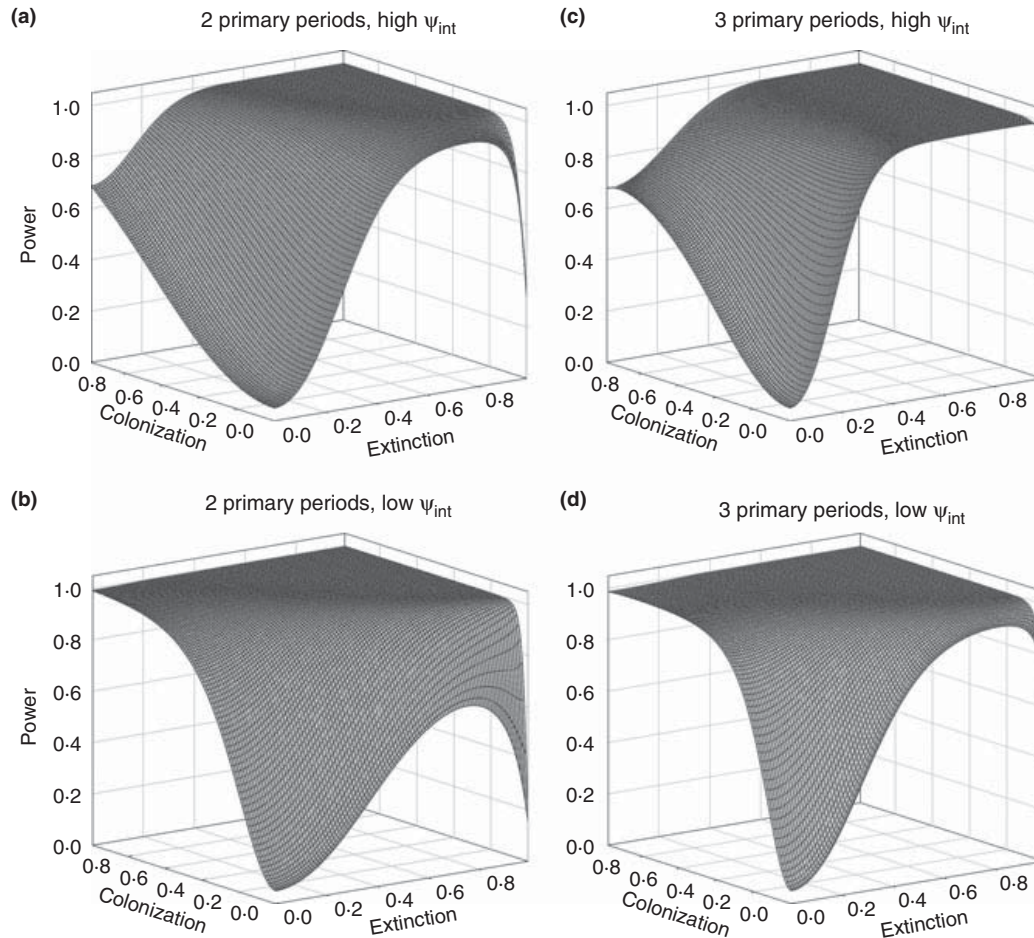
Fig. 4. The power of a likelihood-ratio test to detect changes in site occupancy between primary sampling periods, allowing estimates of detection probability, p, to vary between primary sampling periods. High $\psi_{int} = 0.70$ and low $\psi_{int} = 0.30$. All calculations were made assuming 150 sites and $p = 0.70$.

occupancy should be done with explicit emphasis on the problem of closure. Populations may often be open during time periods typically considered closed by investigators, which can result in biased estimates of occupancy. Consistently high support for open RD models with both the riparian and Hubbard Brook data sets highlights that habitats may frequently be open to changes in site occupancy over the breeding seasons of birds, a time period often considered closed (MacKenzie *et al.* 2003).

A lack of closure may occur for several reasons. For example, Betts *et al.* (2008) demonstrated apparent within-breeding-season movement of black-throated blue warblers *Dendrioca caerulescens* along a habitat gradient, where warblers presumably shifted territories as more reliable information about habitat quality became available. Studies using radio-telemetry and/or colour banding have also demonstrated that individuals may shift territories large distances (e.g. > 5 km) after failed breeding attempts (Walk *et al.* 2004; Fletcher, Koford & Seaman 2006) or in response to seasonal fluctuations in food availability (Klemp 2003).

Our simulation study further demonstrates that estimates of occupancy are sensitive to violations of the closure assumption. Our results are consistent with the predictions

of MacKenzie *et al.* (2006), who drew analogy with Kendall's (1999) evaluation of how violations of the closure assumption bias mark–recapture models. MacKenzie *et al.* (2006) predicted how immigration and emigration-only movement was likely to bias estimates of ψ. Our simulations confirmed these predictions and went one step further by additionally evaluating how simultaneous colonization and extinction events bias estimates of ψ, demonstrating that even small amounts of simultaneous colonization and extinction will lead to overestimates of ψ. In practice, movements are likely to be both into and out of sampling units, which was highlighted with both the riparian and Hubbard Brook data sets (Appendix S1).

The strength and direction of bias resulting from violations of closure is likely to be especially problematic when monitoring rare or declining species in a conservation or management context (e.g. Stauffer, Ralph & Miller 2004). For example, simulations demonstrate that bias in ψ from closed models becomes greater as ψ_{int} becomes smaller. Furthermore, because such biases are typically positive, assuming closure when in fact populations are open will result in overestimates of ψ.

The sensitivity of model performance to changes in site occupancy and the likelihood that sites may be open to changes in site occupancy during time-scales typical of many occupancy studies underscore the importance of addressing the closure assumption. We reiterate MacKenzie *et al.*'s (2006) suggestion of conducting replicate surveys as close in time as possible as a means to minimize violations of the closure assumption. We note, however, that while conducting replicate surveys close in time has been proposed as a way to maximize the likelihood of closure, until now the importance of addressing this assumption has not been explicitly addressed.

An advantage of the sampling approach we highlight is the ability to conduct a formal test for closure, complementing the goodness-of-fit test for single-season occupancy models developed by MacKenzie & Bailey (2004). Both approaches enable an assessment of the adequacy of single-season models to describe the population of interest. Additionally, the test for closure we describe enables prospective power analyses. This, in conjunction with MacKenzie & Royle's (2005) recommendations for maximizing precision of parameter estimates with a limited budget, should prove useful in designing occupancy studies.

A potential source of bias in open models could arise if p was to decline between primary sampling periods. Our simulations suggest, however, that open models can generally distinguish between changes in occupancy and detectability, except when ψ_{int} is low and ε is high. If p declines to values approaching zero, neither open nor closed models will provide unbiased estimates of occupancy. In both the riparian and Hubbard Brook data sets, there was no strong evidence that declining p was a source of bias in open models, as BIC weights favoured models with constant p for most species considered.

One potential problem with our approach of treating individual site visits as primary sampling periods is that temporary emigration may be confounded with local extinction or colonization. For example, if a territory were to overlap a survey site, but not be completely contained within that site, what might be inferred as apparent extinction or colonization could simply be an animal still present in its territory, but absent from the survey site. This situation may have occurred for some species considered, especially wide-ranging species. Nichols *et al.*'s (2008) multi-scale occupancy model, which estimates a parameter reflecting the probability a species is available for sampling, conditional on the species occupying the sampling unit, is one potential approach for dealing with temporary emigration. Other specifications for addressing temporary emigration have also been developed and can be applied to occupancy modelling (Kendall, Nichols & Hines 1997; Kery *et al.* 2009). For example, a restricted form of random temporary emigration, assuming $\gamma_t = 1 - \varepsilon_t$ (MacKenzie *et al.* 2006, p. 206), could be easily included into our modelling framework as a model of intermediate complexity between closed and Markovian models, where the likelihood function would be a simplified version of eqn 1.

The problem of temporary emigration could also be addressed more directly in the study design phase by increasing the area sampled and/or potentially sampling at a finer resolution within the area sampled (e.g. using a tighter sampling grid). If temporary emigration from the sampling unit is driving apparent extinction, expanding the area sampled should provide more support for closed models than at smaller spatial scales. This is less of a problem for the riparian data set because most forest patches sampled were small, such that increasing the point-count radius would have resulted in sampling non-forest habitat. Hubbard Brook, however, is characterized by relatively contiguous habitat, which makes the issue of temporary emigration more relevant for this data set. We further analysed the Hubbard Brook data at a 100-m point-count radius to determine if inferences on violations of closure were sensitive to sampling area. Tests for closure at this radius still resulted in rejecting the null hypothesis of closure between primary sampling periods for all 18 species ($P < 0.01$), although BIC weight of closed models increased substantially for the yellow-bellied sapsucker *Sphyrapicus varius* and decreased for white-breasted nuthatch *Sitta carolinensis*. This suggests that, at least for the yellow-bellied sapsucker, a wide-ranging species, temporary emigration may be occurring occasionally.

If temporary emigration was driving apparent colonization and extinction, we should further expect to consistently reject the closure hypothesis for wide-ranging species. However, many species for which the closure hypothesis was supported are wide-ranging, such as the black-headed grosbeak and downy woodpecker. Further, species body mass (a surrogate of territory size, see Bowman (2003) and Appendix S2) was not a significant predictor of BIC model weight for either the riparian or Hubbard Brook data (C.T. Rota, unpublished results). Thus, despite the potential to confound temporary emigration with local extinction-colonization dynamics, our analyses strongly suggest that sites were open to changes in occupancy over the duration of these two studies.

Adequately addressing the closure assumption is critical to drawing valid inference for both ecological questions and conservation problems. Our simulations demonstrate that occupancy models are sensitive to violations of the closure assumption and will tend to overestimate the probability of site occupancy when closure is violated. Further, this bias is dependent on the amount of local colonization and extinction. These parameters are unlikely to remain constant within or among seasons, making inference even on relative occupancy from closed models problematic. We recommend explicitly addressing the assumption of closure whenever possible, both in sampling designs that include repeated sampling over a time-scale (hours to days) that will maximize the likelihood of closure, and by adopting designs which will allow formal testing for closure. Adequately addressing assumptions is an essential part of any modelling process, and an improved focus on the closure assumption will play a crucial role for managers and scientists alike in providing meaningful estimates of occupancy.

Acknowledgements

References

Azuma, D.L., Baldwin, J.A. & Noon, B.R. (1990) *Estimating the Occupancy of Spotted Owl Habitat Areas by Sampling and Adjusting for Bias*. General Technical Report PSW 124, US Department of Agriculture, Forest Service, Pacific Southwest Research Station, Berkeley.

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

Ball, L.C., Doherty, P.F. & McDonald, M.W. (2005) An occupancy modeling approach to evaluating a Palm Springs ground squirrel habitat model. *Journal of Wildlife Management*, **69**, 894–904.

Betts, M.G., Rodenhouse, N.L., Sillett, T.S., Doran, P.J. & Holmes, R.T. (2008) Dynamic occupancy models reveal within-breeding season movement up a habitat quality gradient by a migratory songbird. *Ecography*, **31**, 592–600.

Bowman, J. (2003) Is dispersal distance of birds proportional to territory size? *Canadian Journal of Zoology*, **81**, 195–202.

Fletcher Jr, R.J. & Hutto, R.L. (2008) Partitioning the multi-scale effects of human activity on the occurrence of riparian forest birds. *Landscape Ecology*, **23**, 727–739.

Fletcher, R.J., Koford, R.R. & Seaman, D.A. (2006) Critical demographic parameters for declining songbirds breeding in restored grasslands. *Journal of Wildlife Management*, **70**, 145–157.

Geissler, P.H. & Fuller, M.R. (1987) Estimation of the proportion of an area occupied by an animal species. *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, **1986**, 533–538.

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

Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overton, J.M.C., Aspinall, R. & Hastie, T. (2006) Making better biogeographical predictions of species' distributions. *Journal of Applied Ecology*, **43**, 386–392.

Hanski, I. (1994) A practical model of metapopulation dynamics. *Journal of Animal Ecology*, **63**, 151–162.

Kendall, W.L. (1999) Robustness of closed capture-recapture methods to violations of the closure assumption. *Ecology*, **80**, 2517–2525.

Kendall, W.L., Nichols, J.D. & Hines, J.E. (1997) Estimating temporary emigration using capture-recapture data with Pollock's robust design. *Ecology*, **78**, 563–578.

Kery, M., Royle, J.A., Plattner, M. & Dorazio, R.M. (2009) Species richness and occupancy estimation in communities subject to temporary emigration. *Ecology*, **90**, 1279–1290.

Klemp, S. (2003) Altitudinal dispersal within the breeding season in the grey wagtail *Motacilla cinerea*. *Ibis*, **145**, 509–511.

Link, W.A. & Barker, R.J. (2006) Model weights and the foundations of multi-model inference. *Ecology*, **87**, 2626–2635.

MacKenzie, D.I. & Bailey, L.L. (2004) Assessing the fit of site-occupancy models. *Journal of Agricultural Biological and Environmental Statistics*, **9**, 300–318.

MacKenzie, D.I. & Royle, J.A. (2005) Designing occupancy studies: general advice and allocating survey effort. *Journal of Applied Ecology*, **42**, 1105–1114.

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.

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., Royle, J.A., Pollock, K.H., Bailey, L.L. & Hines, J.E. (2006) *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence*. Elsevier, Amsterdam.

Marsh, D.M. & Trenham, P.C. (2008) Current trends in plant and animal population monitoring. *Conservation Biology*, **22**, 647–655.

Martin, T.G., Wintle, B.A., Rhodes, J.R., Kuhnert, P.M., Field, S.A., Low-Choy, S.J., Tyre, A.J. & Possingham, H.P. (2005) Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecology Letters*, **8**, 1235–1246.

Moilanen, A. (2002) Implications of empirical data quality to metapopulation model parameter estimation and application. *Oikos*, **96**, 516–530.

Nichols, J.D., Bailey, L.L., O'Connell, A.F., Talancy, N.W., Grant, E.H.C., Gilbert, A.T., Annand, E.M., Husband, T.P. & Hines, J.E. (2008) Multi-scale occupancy estimation and modelling using multiple detection methods. *Journal of Applied Ecology*, **45**, 1321–1329.

Pollock, K.H. (1982) A capture-recapture design robust to unequal probability of capture. *Journal of Wildlife Management*, **46**, 752–757.

Royle, J.A. & Dorazio, R.M. (2008) *Hierarchical Modeling and Inference in Ecology*. Academic Press, Amsterdam.

Self, S.G. & Liang, K.Y. (1987) Asymptotic properties of maximum-likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, **82**, 605–610.

Stauffer, H.B., Ralph, C.J. & Miller, S.L. (2004) Ranking habitat for marbled murrelets: new conservation approach for species with uncertain detection. *Ecological Applications*, **14**, 1374–1383.

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

Walk, J.W., Wentworth, K., Kershner, E.L., Bollinger, E.K. & Warner, R.E. (2004) Renesting decisions and annual fecundity of female dickcissels (*Spiza americana*) in Illinois. *Auk*, **121**, 1250–1261.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Likelihood-ratio tests and power calculations.

Appendix S2. Scientific names, body mass and estimates for each species.

Appendix S3. R code for occupancy models.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.