

## SPECIAL ISSUE. MODELLING DEMOGRAPHIC PROCESSES IN MARKED POPULATIONS: PROCEEDINGS OF THE EURING 2013 ANALYTICAL MEETING

# Advances and applications of occupancy models

Larissa L. Bailey<sup>1\*</sup>, Darryl I. MacKenzie<sup>2</sup> and James D. Nichols<sup>3</sup>

<sup>1</sup>Department of Fish, Wildlife, and Conservation Biology, Colorado State University, 1474 Campus Delivery, Fort Collins, CO 80523, USA; <sup>2</sup>Proteus Wildlife Research Consultants, P.O. Box 5193, Dunedin 9058, New Zealand; and <sup>3</sup>U.S. Geological Survey, Patuxent Wildlife Research Center, 12100 Beech Forest Rd, Laurel, MD 20708, USA

### Summary

1. The past decade has seen an explosion in the development and application of models aimed at estimating species occurrence and occupancy dynamics while accounting for possible non-detection or species misidentification.
2. We discuss some recent occupancy estimation methods and the biological systems that motivated their development. Collectively, these models offer tremendous flexibility, but simultaneously place added demands on the investigator.
3. Unlike many mark–recapture scenarios, investigators utilizing occupancy models have the ability, and responsibility, to define their sample units (i.e. sites), replicate sampling occasions, time period over which species occurrence is assumed to be static and even the criteria that constitute ‘detection’ of a target species. Subsequent biological inference and interpretation of model parameters depend on these definitions and the ability to meet model assumptions.
4. We demonstrate the relevance of these definitions by highlighting applications from a single biological system (an amphibian–pathogen system) and discuss situations where the use of occupancy models has been criticized. Finally, we use these applications to suggest future research and model development.

**Key-words:** amphibian disease, *Batrachochytrium dendrobatidis*, detection probability, dynamic multistate occurrence models, false positive, misidentification, multiscale, study design

### Introduction

Since the seminal work by MacKenzie *et al.* (2002, 2003), there has been an explosion in the development and application of models aimed at estimating species occurrence and occupancy dynamics while accounting for possible non-detection or species misidentification (MacKenzie *et al.* 2006, 2009; Miller *et al.* 2011). Since 2002, well over 1000 papers have cited occupancy models (Google Scholar) with a plethora of studies investigating ecological questions and processes such as species distribution modelling (e.g. Royle, Nichols & Kéry 2005; Kéry, Guillera-Aroita & Lahoz-Monhort 2013), habitat relationships (e.g. Ball, Doherty & McDonald 2005), metapopulation dynamics (e.g. Ferraz *et al.* 2007), invasive species dynamics (e.g. Bled, Royle & Cam 2011; Yackulic *et al.* 2012), multispecies relationships (e.g. competition or predation, MacKenzie *et al.* 2006; Miller *et al.* 2012a) and community dynamics (Zipkin, Dewan & Royle 2009). The studies have involved numerous vertebrate taxa as well as applications to plants (e.g. Kéry 2004), invertebrates (e.g. Govindan, Kéry & Swihart 2012) and pathogens (e.g. Gomez-Dias *et al.* 2010). Uses of

occupancy models have also extended to applications in human medicine, palaeontology (taxonomic ranges based on fossil data, Liow 2013) and even political science (probabilities of incidents reflecting political unrest).

Various extensions of the original static and dynamic models have been proposed to accommodate multiple occupied states (Royle 2004; Royle & Link 2005; Nichols *et al.* 2007; MacKenzie *et al.* 2009), estimate community-level metrics and dynamics (Dorazio & Royle 2005; Kéry & Royle 2009), simultaneously model habitat and occupancy dynamics (Martin *et al.* 2010; MacKenzie *et al.* 2011; Miller *et al.* 2012a), estimate species occurrence at multiple spatial or temporal scales (Nichols *et al.* 2008; Kendall 2009; McClintock *et al.* 2010b; Mordecai *et al.* 2011; Pavlacky *et al.* 2012), and model occupancy dynamics as a function of the occupancy states of nearby (neighbouring) sites (Royle & Dorazio 2008; Bled, Royle & Cam 2011; Yackulic *et al.* 2012). Other model development has been aimed at relaxing model assumptions by allowing heterogeneous detection probabilities (e.g. MacKenzie *et al.* 2006; Royle 2006), including abundance-induced heterogeneity in detection probability (Royle & Nichols 2003; Royle 2004; Royle & Dorazio 2008), dealing with lack of independence among repeated detection surveys at a sampling unit (Nichols *et al.* 2008; Hines *et al.* 2010; Guillera-Aroita 2011), accommodating

\*Correspondence author. E-mail: llbailey@colostate.edu

misidentification or false-positive detections (Miller *et al.* 2011, 2013), and developing methods to address violations of the closure assumption (Rota *et al.* 2009; Kendall *et al.* 2013).

Occurrence data have the advantage of being relatively easy to collect, and the availability of various free software packages has contributed to proliferation of the use of occupancy models (e.g. Program PRESENCE, Hines 2006; Program MARK, White & Burnham 1999; the R package *unmarked*, Fiske & Chandler 2011; OpenBUGS, Lunn *et al.* 2009). Still, the ability to draw strong biological inference from occurrence data depends upon the clear articulation of study objectives and an appropriate sampling design. Investigators utilizing occupancy models have the ability, and responsibility, to define the following terms based on their biological questions, logistical constraints and other study specifics: sample units (i.e. sites), the time period over which occurrence is assumed to be static, replicate surveys and even the criteria that constitute 'detection'. Several recent papers have been critical of the use of occupancy models when the above terms are ambiguous (Efford & Dawson 2012) or defined in a manner that leads to violation of critical model assumptions (e.g. Kendall & White 2009; Guillera-Aroita 2011). While the importance of study design has always been emphasized in the occupancy arena (MacKenzie & Royle 2005; MacKenzie *et al.* 2006; Bailey *et al.* 2007; Guillera-Aroita & Lahoz-Mohort 2012), most occupancy design papers concentrate on developing cost-efficient designs for a generic application (e.g. MacKenzie & Royle 2005; Guillera-Aroita, Ridout & Morgan 2010) or evaluate the relative merits of different detection methods (e.g. Nichols *et al.* 2008). To provide general recommendations on appropriate sample size and optimal allocation of effort among sites and surveys, these papers have necessarily assumed that investigators appropriately define key components to address their biological objectives. Specific recommendations about how to define these key components are difficult to provide because objectives, logistical constraints and other key determinants of study design differ among studies. Such factors determine how well the study system corresponds to model assumptions and thus the strength of inference provided by that application of occupancy models.

In this paper, we discuss some of the newer occupancy models and the biological systems that motivated their development. Collectively, these models offer tremendous flexibility and exciting new ways for practitioners to address biological questions related to species occurrence. We emphasize that biological inference and interpretation of model parameters depend upon the study system and the ability to meet model assumptions. To illustrate the importance of defining key occupancy components, we show how these definitions can vary within a given biological system based on different study objectives. A clear objective that specifies exactly what biological aspect of the system is being represented by 'occupancy' should lead naturally to reasonable options for elements of the field design (e.g. sample unit, season length, etc.). This, in turn, will lead to analyses that yield strong, defensible biological inferences about the system of interest. Failure to specify a clear objective will result in weaker inferences.

## Occupancy models: key components and assumptions

Occupancy is usually defined as the probability that the focal taxon occupies, or uses, a sample unit during a specified period of time during which the occupancy state is assumed to be static. A typical occupancy study design involves identifying the complete set of sample units of interest (of size  $S$ ) and selecting a sample (of size  $s$ ) in a manner that allows investigators to generalize conclusions based upon the sample to the specified population of units. During designated points in time that relate to the time-scale at which occupancy states are likely to change, the  $s$  units are repeatedly surveyed within a relatively short time period (during which the occupancy state at each unit is static), and the observed occupancy state is recorded during each survey of each unit, or 'site'. This design clearly resembles Pollock's robust design in mark-recapture models (Pollock 1982; Kendall, Pollock & Brownie 1995), but within the occupancy literature, primary periods are often referred to as 'seasons' and secondary sessions as 'surveys or visits'. Within each season, the occupancy state of each unit does not change (i.e. closure assumption); hence, the repeated surveys provide multiple opportunities to observe the true occupancy state for a given season. Between seasons, the occupancy state may change at the sites: occupied sites may become unoccupied (i.e. local extinction) and unoccupied sites may be colonized. Model parameters under this simple dynamic model (MacKenzie *et al.* 2003) include the following:

$\psi_{i1}$  = the probability that unit  $i$  is occupied by the target species during the first season.

$p_{ijt}$  = the probability of detecting the species at an occupied unit  $i$  during the  $j$ th independent survey of the site during season  $t$ .

$\epsilon_{it}$  = the probability that occupied unit  $i$  in season  $t$  becomes unoccupied in season  $t + 1$  (local extinction).

$\gamma_{it}$  = the probability that an unoccupied unit  $i$  in season  $t$  is occupied by the target species in season  $t + 1$  (colonization).

Model likelihoods for the data can be developed that explicitly incorporate the biological and sampling processes either by accounting for all possibilities when there is ambiguity in the true occupancy state (due to imperfect detection, i.e. integrating across the occupancy states), or using a hierarchical modelling approach. Parameter estimates may be obtained using maximum likelihood or Bayesian methods of inference.

There are several critical assumptions for the model described above, including (i) the closure assumption mentioned above; (ii) the probability of initial occupancy (season 1,  $\psi_{i1}$ ) and subsequent vital rate parameters ( $\epsilon_{it}$ ,  $\gamma_{it}$ ) are constant among sites, or differences are modelled using covariates (e.g. usually via a logit link); (iii) no unmodelled heterogeneity in detection probabilities; (iv) survey outcomes are independent of one another; and (v) species are not misidentified or falsely detected when a site is unoccupied (MacKenzie *et al.* 2006). If these assumptions are not met, estimators may be biased, precision may be overstated, and inferences about factors influencing model parameters may be incorrect. For greater detail

on the basic modelling approaches, readers should see MacKenzie *et al.* (2002, 2003, 2006).

### Motivation for new model development

During the past decade, several new models have been developed to expand upon the basic dynamic model or relax its associated assumptions. In this section, we discuss some of these models and refer readers to other extensions presented in this session.

#### BEYOND TWO STATES

Soon after the publication of the initial occupancy models, there was a desire to extend models to include more than one occupied state. There are many biological systems where it is advantageous to classify different subcategories of species occurrence (e.g. breeding/non-breeding or relative abundance classes) or simultaneously model the joint dynamics of habitat and species occurrence. Motivated by these biological systems, several dynamic multistate models have been developed to provide inferences about the probabilities of sites being in any one of the occupancy states and making dynamic transitions among them. These models allow for ambiguity not only in the presence or absence of the species (for example) from the field observations, but also in assignment of the correct subcategory. We begin by describing a general model and then highlight the flexibility of this class of model using a diverse group of innovative biological applications.

Motivated by the case of anuran call index data, Royle (2004) and Royle & Link (2005) developed models for assessing patterns in multiple occurrence states at a single point in time. Nichols *et al.* (2007) developed a reparameterized version

these cases, the observed occupancy states are defined hierarchically, such that the lowest observed state (non-detection) has the greatest ambiguity about the true occupancy state (all states, occupied and unoccupied, are possible), but there is no uncertainty regarding the true state at a unit where the highest state is observed.

We develop the model first in terms of three possible states (unoccupied and two occupied states) to be consistent with many model applications and then mention how the model can be extended to more states to accommodate other extensions such as habitat–occupancy dynamics and species co-occurrence models.

Let  $\phi^{[m]}$  be the probability that a unit is in occupancy state  $m$ , where  $\sum_m \phi^{[m]} = 1$ . In the case of three mutually exclusive occupancy states, we can write the probability of a unit being unoccupied as  $1 - \phi^{[1]} - \phi^{[2]}$ . Most applications of the multistate occupancy model have used a conditional binomial parameterization where the probability of the higher state is conditional on species occurrence. Here, the probability of occupancy is defined as  $\psi = \phi^{[1]} + \phi^{[2]}$ , and the probability a unit is in state 2 (the highest state) can be written as conditional on occupancy,  $R = \phi^{[2]}|\psi$ . Thus, the initial state probability vector for the first season of sampling can be defined as  $\phi_0 = [1 - \phi^{[1]} - \phi^{[2]} \quad \phi^{[1]} \quad \phi^{[2]}] = [1 - \psi \quad \psi(1 - R) \quad \psi R]$ . A transition probability matrix  $\phi_t$  is used to describe change in the true state of a unit between seasons  $t$  and  $t + 1$ . This matrix can be written in terms of parameters  $\phi_t^{[m,n]}$ , the probability of a unit transitioning from state  $m$  at time  $t$  to state  $n$  at time  $t + 1$  or as the product of two conditional probabilities: for example  $\psi_{t+1}^{[m]}$ , the probability of a unit being occupied at time  $t + 1$ , given that it was in state  $m$  at time  $t$ , and  $R_{t+1}^{[m]}$ , the probability a unit is in the highest state (2) at time  $t + 1$ , given the unit was in state  $m$  at time  $t$  and is occupied at time  $t + 1$ .

$$\Phi_t = \begin{bmatrix} \phi_t^{[0,0]} & \phi_t^{[0,1]} & \phi_t^{[0,2]} \\ \phi_t^{[1,0]} & \phi_t^{[1,1]} & \phi_t^{[1,2]} \\ \phi_t^{[2,0]} & \phi_t^{[2,1]} & \phi_t^{[2,2]} \end{bmatrix} = \begin{bmatrix} 1 - \psi_{t+1}^{[0]} & \psi_{t+1}^{[0]}(1 - R_{t+1}^{[0]}) & \psi_{t+1}^{[0]}R_{t+1}^{[0]} \\ 1 - \psi_{t+1}^{[1]} & \psi_{t+1}^{[1]}(1 - R_{t+1}^{[1]}) & \psi_{t+1}^{[1]}R_{t+1}^{[1]} \\ 1 - \psi_{t+1}^{[2]} & \psi_{t+1}^{[2]}(1 - R_{t+1}^{[2]}) & \psi_{t+1}^{[2]}R_{t+1}^{[2]} \end{bmatrix}$$

of the Royle & Link (2005) model to assess reproductive success at a unit, given species occurrence. MacKenzie *et al.* (2009) provided a general multistate framework, within which these previous developments represented special cases (parameterizations), and then extended these methods to allow estimation of parameters governing the dynamic processes responsible for change in these occupancy states between seasons. For generality, we present both the conditional binomial parameterization and the multinomial parameterization of MacKenzie *et al.* (2009), but we acknowledge that most applications have utilized the conditional form. The conditional binomial approach may be biologically more reasonable when progression from one occupancy state to another is considered as a series of steps (e.g. species is present or absent at a unit and then given presence, breeding or no breeding occurred). Numerically, this parameterization can be more stable, particularly when covariates are incorporated. Importantly, in all

Conditional on the true state of a unit at a given time, detection probabilities are defined for each parameterization (Table 1).

Following MacKenzie *et al.* (2009), the probability of the observed detection histories  $\mathbf{h}$  collected over  $T$  seasons can be determined succinctly using matrix notation, that is,

$$\Pr(\mathbf{h}|\theta) = \Phi_0 \left[ \prod_{t=1}^{T-1} D(\mathbf{p}_{h,t}) \Phi_t \right] \mathbf{p}_{h,T},$$

where  $\mathbf{p}_{h,t}$  is the detection probability vector for the portion of the full detection history observed in season  $t$  and  $D(\mathbf{p}_{h,t})$  is a diagonal matrix with the elements of  $\mathbf{p}_{h,t}$  on the main diagonal (top left to bottom right) and zero elsewhere. Using a matrix formulation, ambiguity in the true occupancy state from the observed data is resolved through the matrix multiplication, which is essentially just the sum of the probabilities of the various possible outcomes. Assuming the detection histories are

**Table 1.** Detection probabilities associated with dynamic multistate occupancy models. In the multinomial parameterization (A),  $p_{i,j}^{[l,m]}$  is the probability of observing a unit in occupancy state  $l$  during survey  $j$  in season  $t$ , given the unit is in true occupancy state  $m$ . In the conditional binomial parameterization (B),  $p_{i,j}^{[m]}$  is the probability of detecting the target species during survey  $j$  in season  $t$ , given the unit is in true state  $m$ , and  $\delta_{i,j}$  is the probability of correctly classifying the true state as state '2', given the species was detected during survey  $j$  in season  $t$

| True state | Observed state                            |                                   |                             |
|------------|---|-----------------------------------|-----------------------------|
|            | 0   | 1                                 | 2                           |
| (A)        |   |                                   |                             |
| 0          | 1   | 0                                 | 0                           |
| 1          | $(1 - p_{i,j}^{[1,1]})$                   | $p_{i,j}^{[1,1]}$                 | 0                           |
| 2          | $(1 - p_{i,j}^{[1,2]} - p_{i,j}^{[2,2]})$ | $p_{i,j}^{[1,2]}$                 | $p_{i,j}^{[2,2]}$           |
| (B)        |   |                                   |                             |
| 0          | 1   | 0                                 | 0                           |
| 1          | $1 - p_{i,j}^{[1]}$                       | $p_{i,j}^{[1]}$                   | 0                           |
| 2          | $1 - p_{i,j}^{[2]}$                       | $p_{i,j}^{[2]}(1 - \delta_{i,j})$ | $p_{i,j}^{[2]}\delta_{i,j}$ |

independent for each unit, the joint probability for the data (and the model likelihood) is

$$\Pr(\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_s | \boldsymbol{\theta}) = L(\boldsymbol{\theta} | \mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_s) = \prod_{i=1}^s \Pr(\mathbf{h}_i | \boldsymbol{\theta})$$

where  $\boldsymbol{\theta}$  denotes all the parameters in the model. Alternatively, the same underlying modelling structure can be developed within a state-space or hierarchical model (MacKenzie *et al.* 2009).

Most avian applications of the above model have focused on evaluating factors influencing occupancy and reproductive success of nesting raptor species. Martin *et al.* (2009) used the method to explore potential negative impacts of recreational activities on Golden Eagles (*Aquila chrysaetos*) in Denali National Park, Alaska, and found that while there was some evidence of reduced colonization (i.e.  $\psi_{t+1}^{[0]}$  was lower at highly accessible sites), conditional reproduction was better modelled as a function of prey abundance. MacKenzie *et al.* (2012) applied the model to potential nesting territories for California Spotted Owls (*Strix occidentalis occidentalis*) in California. This species is relatively long-lived and exhibits high fidelity to nesting territories. In this case, successive territory occupancy should be correlated with adult survival, and probability of reproductive success should be highly correlated with *per capita* reproductive rates, making it possible to investigate population dynamics without marking individuals. A similar multievent approach was also taken by Lorentzen, Choquet and Steen (2012) to estimate survival and hatching success at occupied nests of colonial seabird species.

Interestingly, several authors have applied the model to estimate habitat and occupancy dynamics by redefining the true states as unsuitable habitat (and thus unoccupied), suitable habitat occupied by no or only few individuals of the target species or suitable habitat occupied by some or many individuals of the target species. Such an approach was applied to study

the use of water holes by African elephants (*Loxodonta africana*) in Hwange National Park, Zimbabwe (Martin *et al.* 2010), and the occurrence of larvae of a suite of plains fishes in the Arikaree River, eastern Colorado, USA (Falke *et al.* 2012). The aim of these studies was to examine how habitat suitability and factors that affect habitat suitability (e.g. rainfall, snowpack) influence the distribution and abundance of target species. Actively manipulating the availability of surface water (habitat) is one option that managers have to influence these species' distributions, and the resulting parameter estimates can be used to predict responses to potential actions under differing environmental conditions to identify optimal management decisions.

Applications such as these prompted the development of more flexible models to separate the fundamental components of habitat and species occurrence dynamics to better understand ecological processes. MacKenzie *et al.* (2011) developed a model that permits variable levels of species occurrence probabilities among multiple habitat types, where the species occurrence may also influence the habitat dynamics (e.g. overgrazing). Likewise, Miller *et al.* (2012a) conceptualized a model that included combinations of habitat (suitable/unsuitable) and predator and prey occurrence. Both papers modelled the dynamics of target amphibian species that rely on ephemeral habitats for long-term persistence, hence the emphasis on modelling habitat and species dynamics simultaneously. These studies provide examples of the flexibility of occupancy models, specifically investigating multiple simultaneous effects on focal species dynamics. At the same time, multiple effects necessitate the development of complex models and require that investigators carefully consider only those models that are relevant for their biological system. We will return to these considerations in later sections.

#### FALSE-POSITIVE DETECTIONS

Arguably the most vital assumption of the occupancy models mentioned thus far is that species are not misidentified or falsely detected when a site is unoccupied. While the possibility of false-positive errors had been acknowledged in various studies (e.g. Simons *et al.* 2007; Shea *et al.* 2011), the problem had been largely ignored in model development until recently (but see Royle & Link 2006). A series of experimentally based papers on birds and anurans highlighted the pervasiveness of the problem in aural detections, noting that false-positive errors existed for nearly all species and observers (Simons *et al.* 2007; McClintock *et al.* 2010a; Miller *et al.* 2012b). Moreover, there was limited ability to reduce these errors with additional, targeted training (Miller *et al.* 2012b). Using standard dynamic occupancy models, low levels of false-positive errors (< 5% of all detections) caused severe overestimation of site occupancy, colonization and local extinction probabilities, as well as spurious relationships between these parameters and explanatory variables (Royle & Link 2006; McClintock *et al.* 2010a; Miller *et al.* 2011, 2013). It should also be noted that biases introduced by species misidentification are not limited



to analyses that account for imperfect detection. If a species may be misidentified, but is detected perfectly at a sample unit otherwise, 'false presences' will result and occupancy estimates are likely biased.

Miller *et al.* (2011, 2013) developed models that accommodate possible false-positive detections, provided a subset of the detections is certain (i.e. a species may be present and not detected, but detections have no false-positive errors; also see Hanks, Hooten & Baker (2011) for a similar Bayesian hierarchical model). The models resemble the multistate models described above and are analogous to the multievent models employed in mark-recapture studies to deal with state uncertainty (Pradel 2005). Initial occupancy and rate parameters are identical to those of the multistate model, where the true occupancy state of a unit  $i$  in season  $t$ ,  $m_{it}$ , is one of  $K$  discrete occupancy states. In the simple case of only two true states (occupied,  $m_{it} = 1$ , or unoccupied  $m_{it} = 0$ ), possible observations on survey  $j$  are non-detection (denoted  $y_{ij} = 0$ ), uncertain detection (meaning detection with the possibility of a false positive;  $y_{ij} = 1$ ) and certain detection ( $y_{ij} = 2$ ). The expected probability of recording an observed state  $y$ , given the true state  $m$  is given in Table 2. Notice that if a detection is considered 'certain' ( $y = 2$ ), the unit is assumed to be occupied (i.e. if a detection is considered 'certain', the species is detected without error). In addition to this model permitting both types of detections at any survey, Miller *et al.* (2011) also developed a complementary model for cases where a subset of sample units is surveyed by different methods on different sample occasions, with one method admitting possible false positives and the other method being certain (detections included no false positives).

False-positive models are relatively new and have only been applied to the ecological studies that helped motivate their development, namely anuran studies that rely heavily on aural detections (Miller *et al.* 2011) and a study investigating occupancy dynamics of wolf packs in Montana by combining hunter observations and radiotelemetry information (Miller *et al.* 2013). We believe that any study that relies on indirect animal detection, such as animal sign (e.g. tracks, scats Karanth *et al.* 2011; Molinari-Jobin *et al.* 2012) or interviews of local experts (e.g. Zeller *et al.* 2011), will benefit from these new models. Additionally, the models could be applied in studies that utilize

computer algorithms (e.g. Waddle, Thigpen & Glorioso 2009) or laboratory assays (e.g. McClintock *et al.* 2010b) to determine species identification from survey results. Given the severe bias that can result from ignoring false-positive detections, we hope that any study that may suspect such errors would use these models to formally test whether  $p_{t,j}^{[10]} = 0$ . Utilizing a combination of design modifications to lower the prevalence of false positives and model-based approaches to deal with problems that remain should reduce the bias caused by false-positive detections.

## MULTISCALE MODELS

The work described above details methods that allow for the expansion of the number of occupied states, but another arena of rapid development has focused on expanding the number of hierarchical scales of species occurrence. Motivation for these models includes relaxing model assumptions, such as independence among surveys and closure, and differentiating between species occurrence at local and larger scales. Again, we begin by describing a single-season multiscale model and then highlight variations of this model theme using a diverse group of innovative biological applications.

Early multiscale occupancy models were developed to address lack of independence, or correlation, among surveys (Nichols *et al.* 2008; Hines *et al.* 2010). Often multiple detection devices are deployed at the same location within a sample unit to detect multiple species, or individuals of various life-history phases for a given species, or to compare the efficiency of multiple detection devices (see citations within Nichols *et al.* 2008). If detections from each device are used as surveys, lack of independence may exist if individual animals detected by one device are more likely to be detected by another device (i.e. detections among surveys are not independent). Nichols *et al.* (2008) exploited this dependence to permit inference about species occurrence at two hierarchical scales, the small scale of the location at which sampling devices were deployed and the larger scale of the sample unit within which the devices were located. The basic sampling design is identical to the general framework described above, but  $L$  different surveys (e.g. detection devices, observers or timed observations) are collocated in each sampled unit and sampled at  $T$  occasions or subunits. The occupancy state of the unit is assumed static over this time period, but the species local availability (e.g. presence at the specific location of the detection devices) may change over time or subunits. Model parameters under this model include the following:

$\psi_i$  = the probability that sample unit  $i$  is occupied by the target species

$\theta_{it}$  = the probability the species is locally present (available for detection) at occasion or subunit  $t$ , given the unit  $i$  is occupied.

$p_{itj}$  = the probability of detecting the species with survey  $j$ , given that it is locally present at occasion or subunit  $t$ .

The two occupancy parameters,  $\psi_i$  and  $\theta_{it}$ , permit the modelling of occupancy at two different scales (spatial or temporal):  $\psi_i$  corresponds to species occurrence at the larger scale,

**Table 2.** The probability of recording an observed state  $y$ , given the true state  $m$ , using occupancy models that allow for 'false-positive' detections.  $p_{t,j}^{[10]}$  is the probability of incorrectly detecting the species during survey  $j$  of season  $t$  at an unoccupied unit (possible only for uncertain detections),  $p_{t,j}^{[11]}$  is the probability of detecting the species during survey  $j$  of season  $t$  at an occupied unit, and  $b_{t,j}$  is the probability that a detection is classified as certain during survey  $j$  of season  $t$ , given that the unit is occupied and the species is detected

| True state | Observed state, $y$  |                               |                         |
|------------|----------------------|-------------------------------|-------------------------|
|            | 0                    | 1                             | 2                       |
| 0          | $1 - p_{t,j}^{[10]}$ | $p_{t,j}^{[10]}$              | 0                       |
| 1          | $1 - p_{t,j}^{[11]}$ | $p_{t,j}^{[11]}(1 - b_{t,j})$ | $p_{t,j}^{[11]}b_{t,j}$ |

while  $\theta_{it}$  refers to the presence of the target species at the local scale, conditional on species presence in the sample unit (the larger scale). The product  $\psi_i \theta_{it}$  represents the unconditional probability of small-scale occupancy, indicating presence of individual(s) of the species at the local spatial or temporal scale, and  $\psi_i (1 - \theta_{it})$  represents the probability that the species is present in the unit, but unavailable for detection at occasion or subunit  $t$ .

Pavlacky *et al.* (2012) employed this approach to estimate occupancy and local availability for two avian species thought to differ in characteristics that make them rare. They used a common sampling approach where sample units (1 km<sup>2</sup> plots) were chosen in a probabilistic manner and then multiple point count stations were systematically placed at equal distances (250 m) from one another within chosen units. Each fixed-radius (125 m) point count station was surveyed using a time-to-detection method during the breeding season. Lark sparrow (*Chondestes grammacus*) were fairly scarce among sample units ( $\hat{\psi} = 0.2$ ), but were locally common when they occurred ( $\hat{\theta} = 0.35$ ). Conversely, brown creepers (*Certhia Americana*) occupied more units across the landscape ( $\hat{\psi} = 0.3$ ), but were locally rare ( $\hat{\theta} = 0.1$ ), identifying this species as more susceptible to future declines at the regional scale. In a similar example, Mordecai *et al.* (2011) investigated factors influencing occurrence and temporal availability of Louisiana waterthrush (*Seiurus motacilla*) at point count locations along stream transects in West Virginia, USA. While Pavlacky *et al.* (2012) assumed conditional independence among their point count stations, Mordecai *et al.* (2011) accounted for possible spatial dependence among point counts by including a random intercept for aggregates (transects) of point count stations.

A severe form of spatial dependence may occur for species that are detected along transects or trails, when units are only surveyed once ( $L = 1$ ) and trail segments are used as spatial replicates. Such designs are often employed for large carnivore species, such as tigers, and two different modelling approaches have been proposed to deal with this type of problem (Hines *et al.* 2010; Guillera-Arroita *et al.* 2011). One approach describes the detection process as a continuous point process, where detections occur randomly along a continuous axis (Poisson process) or potential clustering in detections are accounted for via a Markov modulating Poisson process (Guillera-Arroita *et al.* 2011). Another approach is to discretize the trail or transect into spatial subunits of equal length and then model spatial dependence as a first-order Markov process by defining two parameters for local occurrence (Hines *et al.* 2010):

$\theta_{it}$  = the probability the species is present (available) at subunit  $t$  within an occupied unit  $i$ , given the species was not present at the previous subunit.

$\theta'_{it}$  = the probability the species is present (available) at subunit  $t$  within an occupied unit  $i$ , given the species was present at the previous subunit.

Detection histories for each sampled unit consist of detection–nondetection data from each successive spatial subunit, for example,  $h_i = 01011$  denotes a unit where 5 successive spatial subunits were surveyed. Assuming constancy in model

parameters among units, the probability statement associated with this history is  $\Pr(h_i = 01011) = \psi[(1 - \theta_1)\theta_2 + \theta_1(1 - p_1)\theta'_2]p_2[(1 - \theta'_3)\theta_4 + \theta'_3(1 - p_3)\theta'_4]p_4\theta'_5p_5$ . Note the terms in square brackets account for the ambiguity associated with the non-detection of the species in the first and third subunits. Every detection history can be modelled in this manner, and the likelihood under this model can be expressed as follows:  $L(\psi, \theta, \theta', p | h_1, h_2, \dots, h_s) = \prod_{i=1}^s \Pr(h_i)$ . Covariates can be used to model variation in any of the model parameters, but with only a single survey at each spatial subunit, there is limited ability to distinguish between factors influencing local occurrence (availability) and those influencing the conditional detection probability. To date, most applications have assumed that local occurrence parameters are constant over subunits and modelled detection probability as a function of covariates (e.g. Karanth *et al.* 2011). If repeated surveys are conducted at each subunit, better differentiation of factors influencing local occurrence and detection probability is possible, while still accommodating spatial dependence in occupancy (availability) among spatial subunits (J.E. Hines and L.L. Bailey, unpublished data).

The applications mentioned above have all involved surveys of spatial subunits within a larger unit, but it is easy to imagine the same general framework for temporal subunits or occasions within a longer time period of interest. For example, previous avian studies often involve a single visit to each unit, with detections recorded in multiple, successive time periods (i.e. species recorded in 3.5-min periods during a 15-min point count). Even if a unit is visited multiple times within a season, the closure assumption is violated by the non-random occurrence (availability) of a species during the season. Investigators have addressed this issue by (i) creating extra replication (surveys) for each occasion  $t$ , then applying the standard dynamic occupancy model (MacKenzie *et al.* 2003) to directly estimate temporal availability,  $\theta_{it}$ , and deriving a 'large-scale' parameter corresponding to the longer time period of interest (Rota *et al.* 2009), or (ii) modelling species arrival and departure times directly for species with staggered entry and exit times during the period of interest (Kendall *et al.* 2013).

Like the multistate models described in a previous section, multiscale models offer flexibility that is often necessary to address important model assumptions, but they are not meant as a fix for poor study design. Their application is essential to control for biases in certain biological and sampling scenarios, but the additional model complexity may lead to poorer precision and weaker inference.

### Model flexibility and study design: disease system example

Occurrence data have the advantage of being relatively easy to collect, and historic records can often be converted to detection–nondetection data. The availability of numerous data sets and a variety of flexible occupancy models have led to many occupancy-based papers in the literature. For many of these applications in which the underlying ecological and data collection processes were well approximated by occupancy

models, reasonable inferences were obtained. However, when data collection and study system do not correspond well to the processes for which occupancy models were developed, reasonable inferences are not necessarily expected. It is the responsibility of investigators utilizing any model based on detection–nondetection data to clearly define the following terms as applied to their study objectives: sample units (i.e. sites), the time period over which occurrence is assumed to be static, replicate surveys and even the criteria that constitute ‘detection’. Still, many practitioners overlook the importance of defining key terms (e.g. site, survey, season) with respect to their biological question(s) and focus instead on the practical trade-offs related to the optimal number of surveys per site. Most occupancy design papers have followed suit, concentrating on developing cost-efficient designs for a generic application (e.g. MacKenzie & Royle 2005). But while there are some aspects of study design that can be usefully treated in a generic manner, other aspects require a tailoring of design to ecological and sampling specifics. Such specifics typically involve first specifying study objectives, which lead to specific definitions of key model terms. Conditional on these objectives and definitions, appropriate data are collected and model(s) are developed to correspond to the underlying processes of interest.

In the following section, we focus on a single biological system (a host–pathogen system) to illustrate how different biological hypotheses (objectives) result in dramatically different study designs. In each study, we emphasize the definition of key occupancy components and the associated model assumptions that are most relevant to the biological questions being addressed.

#### DISEASE SYSTEM: BACKGROUND

Many amphibian declines world-wide have been attributed to the emerging infectious disease chytridiomycosis, caused by the fungal pathogen *Batrachochytrium dendrobatidis* (hereafter *Bd*, e.g. Berger *et al.* 1998; Muths *et al.* 2003). *Bd* is transmitted between individuals and the environment via an aquatic flagellated zoospore (Berger *et al.* 2005). When the load of zoospores on an individual is high enough, it can alter electrolyte transport across the epidermis, disrupting ion homeostasis, and lead to cardiac arrest (Voyles *et al.* 2009). Susceptibility of amphibians to chytridiomycosis is variable among species; vulnerable species often decline rapidly, while resistant species may function as a reservoir for the pathogen.

Despite numerous papers focusing on amphibian–*Bd* interactions, few have considered imperfect detection of either the host or the pathogen (but see Adams *et al.* 2010; Miller *et al.* 2012c). The following occupancy-based examples are a mixture of published works specific to the *Bd* disease system and proposed designs, some motivated by different ecological systems.

#### PATHOGEN PREVALENCE IN A SINGLE HOST POPULATION

Often the most fundamental parameter in disease studies is prevalence, the proportion of infected individuals in a defined

population of organisms (i.e. disease frequency). Studies of *Bd* dynamics have focused on both prevalence and the infection intensity, defined as the abundance of *Bd* found on infected individuals (e.g. Briggs, Knapp & Vredenburg 2010; Miller *et al.* 2012c). In these studies, the ‘area of interest’ is usually a single amphibian population, where a subset of individuals (sample units) is randomly selected (presumably). Multiple surveys are obtained from all or a subset of captured individuals, and the manner in which these surveys are collected define the time period over which prevalence is assumed to be static. Typically, *Bd* is detected on a captured individual by gently rubbing the surface of the skin with a cotton swab. Multiple PCR samples (surveys) are prepared from each swab and analysed using quantitative PCR (qPCR) techniques, yielding detection–non-detection information and a quantitative measure of zoospore equivalents for each survey (Hyatt *et al.* 2007). Under this sampling scenario, the resulting estimates of occupancy (prevalence) represent the probability of *Bd* occurrence among individuals in the population and apply to the time period over which individuals were captured, often only a single visit. As with most disease assays, the sensitivity of the PCR is  $<1$  (Hyatt *et al.* 2007), and while most authors acknowledge this fact, they attempt to account for non-detection by simply aggregating results for multiple PCR surveys rather than estimate (and correct for) detection probability (but see Miller *et al.* 2012c).

Several previous studies utilized occupancy models with this type of data to estimate prevalence and address biological questions related to the individual characteristics (e.g. species, life stage) that may influence the probability of pathogen occurrence or detection (Gomez-Dias *et al.* 2010; Cooch *et al.* 2012). These applications rely on the basic assumptions outlined in the previous Occupancy Models section. Many of these assumptions are likely met for the sampling design described above (e.g. closure assumption, no false detections), and others can be addressed by modelling heterogeneity in pathogen occurrence and detection as a function of covariates specific to the individual (unit). However, in many disease systems, the detection of the pathogen is likely a function of the intensity of the pathogen on/in the host. In cases where no index of infection intensity is available, utilizing an approach that models detection probability as a function of the latent distribution of pathogen abundance should reduce the bias in prevalence caused by heterogeneity in *Bd* detection among individuals (Royle & Nichols 2003; Lachish *et al.* 2012). Many *Bd* studies now estimate an index of infection intensity (zoospore equivalents) for each survey. Miller *et al.* (2012c) described two analytical methods to accommodate the relationship between pathogen detection and infection intensity: an ad hoc approach using closed population abundance estimators (Huggins 1991) and a hierarchical Bayesian estimator that extended previous occupancy models to account for observational error in the detection of *Bd* and sampling error in measuring the associated *Bd* zoospore equivalents.

An important assumption in these scenarios/applications is that the sampled units (captured individuals) are a random sample of the population of interest. Studies of other disease

systems have emphasized that estimates of prevalence can be biased if infected and uninfected individuals have different capture probabilities (reviewed in Cooch *et al.* 2012). Establishing whether such detection differences exist among individuals usually requires multiple detections and observations of the disease state for individuals over time and typically utilizes mark–recapture methods (Cooch *et al.* 2012). Such studies are rare for amphibian–*Bd* systems, and none has found differences in individual capture probability associated with *Bd* occurrence, but none has properly accounted for uncertainty in individual infection state (Murray *et al.* 2009; Pilliod *et al.* 2010).

#### ESTIMATING THE PROPORTION OF INFECTED HOST POPULATIONS

Understanding factors that influence a pathogen's distribution and determining when and how it is transmitted among seemingly isolated host populations is a major theme in disease ecology and geographical epidemiology. Numerous studies have found that *Bd* is widely distributed geographically; however, the utility of these studies is limited due to the opportunistic nature of the sampling (Muths, Pedersen & Pedersen 2009). A more robust sampling design might define an 'area of interest' as a collection of perhaps-isolated amphibian populations within a specified region (e.g. amphibian populations at wetlands within a national park). A sample of these populations (units) is drawn in a manner that allows for generalization to the entire collection of amphibian populations. Populations are visited once to capture and swab individuals, often of different species, without replacement: these individual swabs can be viewed as surveys of the unit. Typically in these large-scale studies, a single PCR assay is conducted on each individual swab, or multiple swabs may be pooled, due to budget constraints. Inevitably, the number of surveys (PCR assays) varies among units, though a maximum is usually defined.

Adams *et al.* (2010) used this approach to investigate factors influencing spatial patterns in *Bd* occurrence in amphibian populations in Oregon and northern California, USA. They expected the processes that resulted in the presence of *Bd* in amphibian populations (e.g. variables associated with *Bd* thermal tolerances) would differ from the processes that result in the presence of *Bd* among individuals within an infected population (e.g. species, life-history stage, date). The authors recognized several limitations of applying occupancy models to this design. First, there is a finite number of surveys, determined by the number of amphibians available in the local population, and sampling individuals without replacement creates dependence among surveys. They used simulations to determine that bias resulting from such dependence was negligible for most populations in their study (i.e. for populations >50 individuals with true prevalence values for infected populations > 0.05). Additionally, the probability estimates represent a combination of at least two processes: the probability the pathogen is present on the individual in an infected population (i.e. true prevalence) and the probability of detecting the pathogen using standardized field and laboratory techniques, given it is present

on the individual. While these authors believed that the probability of detecting the pathogen on an infected individual was quite high, the resulting detection probability should be interpreted as an index of prevalence (Miller *et al.* 2012c).

#### INVESTIGATING FACTORS INFLUENCING HOST–PATHOGEN DYNAMICS

Little is known about the long-term dynamics of *Bd* once established in an area of interest. The fungus has low mobility and is considered vulnerable outside of a host (Piotrowski, Annis & Longcore 2004). Recent work has shown differential susceptibility among amphibian species suggesting that some non-target species may function as reservoirs, or vectors, leading to the persistence or spread of *Bd* among habitats and populations (Briggs, Knapp & Vredenburg 2010). Even among susceptible species, it is apparent that *Bd* is *not* invariably lethal and that the pathogen persists in an enzootic state in hosts that survived infection during an initial epizootic (Briggs, Knapp & Vredenburg 2010; Pilliod *et al.* 2010). Understanding amphibian–*Bd* dynamics after an epizootic and determining whether amphibians may persist or recolonize affected areas requires the ability to sample *Bd* at sites that have few or no amphibians.

In this case, a sampling design might define an 'area of interest' as a collection of habitats (e.g. ponds) that may serve as amphibian breeding locations within a specified region. A sample of these habitats (units) is surveyed multiple times to detect both *Bd* and target amphibians during a time period where the occurrence of both pathogen and host is considered static. Captured amphibians can be sampled for *Bd* in the manner described above, but if amphibians are not detected, a water sample could be used to detect *Bd* in the environment (Kirshtein *et al.* 2007). Here, target amphibian detection informs the amphibian state (occupied) at the unit and serves as a survey for the detection of the pathogen. Both swabs and water samples are considered independent surveys of *Bd* at a unit, but the survey methods are likely to have different *Bd* detection probabilities (Kirshtein *et al.* 2007; Hyman & Collins 2012), and these can be accommodated in the modelling.

To our knowledge, no such study exists for an amphibian–*Bd* system, but McClintock *et al.* (2010b) outlined a study design for such a system. In this case, occupancy corresponds to the spatial prevalence of the pathogen across habitats. More interesting are the host–pathogen dynamics that resemble a dynamic species co-occurrence model that may incorporate neighbourhood autologistic effects (McClintock *et al.* 2010b; Yackulic *et al.* 2012, in press). Alternatively, similarities could also be drawn to existing habitat–occupancy dynamic models with two types of habitats: those with and without the pathogen (MacKenzie *et al.* 2011). Relevant biological hypotheses would involve differences in the occurrence and dynamics of the target amphibian species conditional on the habitat type, and importantly, the occurrence of the target species may influence the dynamics of the pathogen (habitat).



## SUMMARY

Each of the studies above addresses different biological questions related to a common host–pathogen system. Accordingly, the definition of a sample unit varies from a single individual to a patch of potential breeding habitat. Investigators have more control over sample unit selection in the latter case, and the inference to all sample units is more defensible. Surveys are defined as individual swabs or groups of swabs, water samples or combinations of these methods. The time period over which these surveys are obtained, ranging from a single visit to an entire breeding season, defines the time period over which the pathogen state, or amphibian and pathogen state, is assumed to be static among units. The longer time periods separating such ‘primary sampling occasions’ (MacKenzie *et al.* 2003) then define the periods to which the occupancy–dynamic rate parameters apply. It is imperative to align the biological hypotheses of the study with a comparable sample design such that selected models are believed to correspond reasonably well to the processes that generated the data; thus, model assumptions are likely to be met.

## Conclusions

The applications presented in this paper are intended to demonstrate the breadth of flexibility possible with current occupancy models. We have chosen to focus on those models that involve a single to a few target species, but we acknowledge that there is a rich literature involving the occurrence of multiple species (e.g. Dorazio & Royle 2005; Kéry & Royle 2009; Zipkin, Dewan & Royle 2009). Ecologists recognize that multiple effects are likely relevant to most studies of occupancy dynamics, and a desire to include these effects motivated the development of the models described here. Cries for even more flexible models to deal with system complexity place added responsibility not only on model developers (biostatisticians), but also on ecologists to clarify and communicate their hypotheses about the dynamics motivating the desire for increased model complexity. We view this added responsibility as a good thing, but simply note that the flexibility of occupancy models requires ecologists to try to restrict the investigated model set to those combinations of effects that represent plausible *a priori* hypotheses.

We recognize that there are relatively few papers that focus on the design of occupancy studies, due in part to the flexibility of existing models. It is difficult, if not impossible, to develop generic recommendations for all aspects of study design. Any attempt to do so would require using language that is so vague (to be inclusive) that it would lose its utility. What is general and conserved is the process that one should go through when designing an occupancy study. Typically, this process involves first specifying study objectives, which then direct definitions of key model terms. Conditional on these objectives and definitions, the process then entails collection of appropriate data and utilization or development of models to correspond to the underlying processes of interest. We hope that our abbreviated demonstration of this process using a single ecological system will assist other practitioners when considering the use of

occupancy models to address hypotheses related to occupancy dynamics.

Many methodological extensions have been developed to address interesting biological questions, but some extensions have emerged from study designs where assumptions of simpler modelling approaches are not met. Such extensions should not, necessarily, be seen as encouraging the particular design used in that study as the best way to address those (or similar) objectives. Whenever possible, one should opt for the simplest and most appropriate study design to address the biological question(s) of interest, instead of relying on model-based solutions to correct for certain aspects of the study design after the data have been collected. In the latter situation, inferences are going to be more model dependent than when potential issues are identified and dealt with during the study design phase.

## References

- Adams, M.J., Chelgren, N.D., Reinitz, D., Cole, R.A., Rachowicz, L.J., Galvan, S. *et al.* (2010) Using occupancy models to understand the distribution of an amphibian pathogen, *Batrachochytrium dendrobatidis*. *Ecological Applications*, **20**, 289–302.
- Bailey, L.L., Hines, J.E., Nichols, J.D. & MacKenzie, D.I. (2007) Sampling design trade-offs in occupancy studies with imperfect detection: examples and software. *Ecological Applications*, **17**, 281–290.
- 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.
- Berger, L., Speare, R., Daszak, P., Green, D.E., Cunningham, A.A., Goggin, C.L. *et al.* (1998) Chytridiomycosis causes amphibian mortality associated with population declines in the rain forests of Australia and Central America. *Proceedings of the National Academy of Sciences*, **95**, 9031–9036.
- Berger, L., Hyatt, A.D., Speare, R. & Longcore, J.E. (2005) Life cycle stages of the amphibian chytrid *Batrachochytrium dendrobatidis*. *Diseases of Aquatic Organisms*, **68**, 51–63.
- Bled, F., Royle, J.A. & Cam, E. (2011) Hierarchical modeling of an invasive spread: the Eurasian Collared-Dove *Streptopelia decaocto* in the United States. *Ecological Applications*, **21**, 290–302.
- Briggs, C.J., Knapp, R.A. & Vredenburg, V.T. (2010) Enzootic and epizootic dynamics of the chytrid fungal pathogen of amphibians. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 9695–9700.
- Cooch, E.G., Conn, P.B., Ellner, S.P., Dodson, A.P. & Pollock, K.H. (2012) Disease dynamics in wild populations: modeling and estimation: a review. *Journal of Ornithology*, **152**, S485–S509.
- Dorazio, R.M. & Royle, J.A. (2005) Estimating size and composition of biological communities by modeling the occurrence of species. *Journal of the American Statistical Association*, **100**, 389–398.
- Efford, M.G. & Dawson, D.K. (2012) Occupancy in continuous habitat. *Ecosphere*, **3**, 32.
- Falke, J.A., Bailey, L.L., Fausch, K.D. & Bestgen, K.R. (2012) Colonization and extinction in dynamic habitats: an occupancy approach for a Great Plains stream fish assemblage. *Ecology*, **93**, 858–867.
- Ferraz, G., Nichols, J.D., Hines, J.E., Stouffer, P.C., Bierregaard, R.O. & Lovejoy, T.E. (2007) A large-scale deforestation experiment: effects of patch area and isolation on Amazon birds. *Science*, **315**, 238–241.
- Fiske, I. & Chandler, R. (2011) Unmarked: an R package for fitting hierarchical models of wildlife occurrence and abundance. *Journal of Statistical Software*, **43**, 1–23.
- Gomez-Dias, E., Doherty, P.F., Duneau, D. & McCoy, K.D. (2010) Cryptic vector divergence masks vector-specific patterns of infection: an example from the marine cycle of Lyme borreliosis. *Evolutionary Applications*, **3**, 391–401.
- Govindan, J.N., Kéry, M. & Swihart, R.K. (2012) Host selection and responses to forest fragmentation in acorn weevils: inferences from dynamic occupancy models. *Oikos*, **121**, 623–633.
- Guillera-Aroita, G. (2011) Impact of sampling with replacement in occupancy studies with spatial replication. *Methods in Ecology and Evolution*, **2**, 401–406.
- Guillera-Aroita, G. & Lahoz-Mohort, J.J. (2012) Designing studies to detect differences in species occupancy: power analyses under imperfect detection. *Methods in Ecology and Evolution*, **3**, 860–869.

- Guillera-Arroita, G., Ridout, M.S. & Morgan, B.J.T. (2010) Design of occupancy studies with imperfect detection. *Methods in Ecology and Evolution*, **1**, 131–139.
- Guillera-Arroita, G., Morgan, B.J.T., Ridout, M.S. & Linkie, M. (2011) Species occupancy modeling for detection data collected along a transect. *Journal of Agricultural, Biological, and Environmental Statistics*, **16**, 301–317.
- Hanks, E.M., Hooten, M.B. & Baker, F.A. (2011) Reconciling multiple data sources to improve accuracy of large-scale prediction of forest disease incidence. *Ecological Applications*, **21**, 1173–1188.
- Hines, J.E. (2006) *PRESENCE3 – Software to estimate patch occupancy and related parameters*. USGS – PWRC. URL: <http://www.mbr-pwrc.usgs.gov/software/presence.html>
- Hines, J.E., Nichols, J.D., Royle, J.A., MacKenzie, D.I., Gopalaswamy, A.M., Kumar, N.S. & Karanth, K.U. (2010) Tigers on trails: occupancy modeling for cluster sampling. *Ecological Applications*, **20**, 1456–1433.
- Huggins, R.M. (1991) Some practical aspects of a conditional likelihood approach to capture experiments. *Biometrics*, **47**, 725–732.
- Hyatt, A.D., Boyle, D.G., Olsen, V., Boyle, D.B., Berger, L., Obendorf, D. *et al.* (2007) Diagnostic assays and sampling protocols for the detection of *Batrachochytrium dendrobatidis*. *Diseases of Aquatic Organisms*, **73**, 175–192.
- Hyman, O.J. & Collins, J.P. (2012) Evaluation of a filtration-based method for detecting *Batrachochytrium dendrobatidis* in natural bodies of water. *Disease of Aquatic Organisms*, **97**, 185–195.
- Karanth, K.U., Gopalaswamy, A.M., Kumar, N.S., Vaidyanathan, S., Nichols, J.D. & MacKenzie, D.I. (2011) Monitoring carnivore populations at landscape scales: occupancy modeling of tigers from sign surveys. *Journal of Applied Ecology*, **48**, 1048–1056.
- Kendall, W.L. (2009) One size does not fit all: adapting mark-recapture and occupancy models for state uncertainty. *Modeling Demographic Processes in Marked Populations* (eds D.L. Thomson, E.G. Cooch & M.J. Conroy), pp. 765–780. Springer, New York.
- Kendall, W.L., Pollock, K.H. & Brownie, C. (1995) A likelihood-based approach to capture-recapture estimation of demographic parameters under the robust design. *Biometrics*, **51**, 293–308.
- Kendall, W.L. & White, G.C. (2009) A cautionary note on trading spatial for temporal sampling in studies of site occupancy. *Journal of Applied Ecology*, **46**, 1182–1188.
- Kendall, W.L., Hines, J.E., Nichols, J.D. & Grant, E.H.C. (2013) Relaxing the closure assumption in occupancy models: staggered arrival and departure times. *Ecology*, **94**, 610–617.
- Kéry, M. (2004) Extinction rate estimates for plant populations in revisitation studies: importance of detectability. *Conservation Biology*, **18**, 570–574.
- Kéry, M., Guillera-Arroita, G. & Lahoz-Monhort, J.J. (2013) Analysing and mapping species range dynamics using occupancy models. *Journal of Biogeography*, **40**, 1463–1474.
- Kéry, M. & Royle, J.A. (2009) Inferences about species richness and community structure using species-specific occupancy models in the national Swiss breeding bird survey MHB. *Modeling Demographic Processes in Marked Populations* (eds D.L. Thomson, E.G. Cooch & M.J. Conroy), pp. 639–656. Springer, New York.
- Kirshstein, J.D., Anderson, C.W., Wood, J.S. & Longcore, J.E. (2007) Quantitative PCR detection of *Batrachochytrium dendrobatidis* DNA from sediments and water. *Disease of Aquatic Organisms*, **77**, 11–15.
- Lachish, S., Gopalaswamy, A.M., Knowles, S.C.L. & Sheldon, B.C. (2012) Site-occupancy modeling as a novel framework for assessing test sensitivity and estimating wildlife disease prevalence from imperfect diagnostic tests. *Methods in Ecology and Evolution*, **3**, 339–348.
- Liow, L.H. (2013) Simultaneous estimation of occupancy and detection probabilities: an illustration using Cincinnati brachiopods. *Paleobiology*, **39**, 193–213.
- Lorentzen, E., Choquet, R. & Steen, H. (2012) Modelling state uncertainty with photo series data for the estimation of breeding success in a cliff-nesting seabird. *Journal of Ornithology*, **152**, S477–S483.
- Lunn, D., Spiegelhalter, D., Thomas, A. & Best, N. (2009) The BUGS project: evolution, critique, and future directions. *Statistics in Medicine*, **28**, 3049–3067.
- 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, A.J. & 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., Hines, J.E. & Bailey, L.L. (2006) *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence*. Elsevier, San Diego, California.
- MacKenzie, D.I., Nichols, J.D., Seamans, M.E. & Guitierrez, R.J. (2009) Modeling species occurrence dynamics with multiple states and imperfect detection. *Ecology*, **90**, 823–835.
- MacKenzie, D.I., Bailey, L.L., Hines, J.E. & Nichols, J.D. (2011) An integrated model of habitat and species occurrence dynamics. *Methods in Ecology and Evolution*, **2**, 612–622.
- MacKenzie, D.I., Seamans, M.E., Guitierrez, R.J. & Nichols, J.D. (2012) Investigating the population dynamics of California spotted owls without marked individuals. *Journal of Ornithology*, **152**, S597–S604.
- Martin, J., McIntyre, C.L., Hines, J.E., Nichols, J.D., Schmutz, J.A. & MacCluskie, M.C. (2009) Dynamic multistate site occupancy models to evaluate hypotheses relevant to conservation of Golden Eagles in Denali National Park, Alaska. *Biological Conservation*, **142**, 2726–2731.
- Martin, J., Chamallé-Jammes, S., Nichols, J.D., Fritz, H., Hines, J.E., Fonnebeck, C.J., MacKenzie, D.I. & Bailey, L.L. (2010) Simultaneous modeling of habitat suitability, occupancy, and relative abundance: African elephants in Zimbabwe. *Ecological Applications*, **20**, 1173–1182.
- McClintock, B.T., Bailey, L.L., Pollock, K.H. & Simons, T.R. (2010a) Unmodeled observation error induces bias when inferring patterns and dynamics of species occurrence based on auditory detections. *Ecology*, **91**, 2446–2454.
- McClintock, B.T., Nichols, J.D., Bailey, L.L., MacKenzie, D.I., Kendall, W.L. & Franklin, A.B. (2010b) Seeking a second opinion: uncertainty in disease ecology. *Ecology Letters*, **13**, 659–674.
- Miller, D.A., Nichols, J.D., McClintock, B.T., Grant, E.H.C., Bailey, L.L. & Weir, L.A. (2011) Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification. *Ecology*, **92**, 1422–1428.
- Miller, D.A.W., Brehme, C.S., Hines, J.E., Nichols, J.D. & Fisher, R.N. (2012a) Joint estimation of habitat dynamics and species interactions: disturbance reduces co-occurrence of non-native predators with an endangered toad. *Journal of Animal Ecology*, **81**, 1288–1297.
- Miller, D.A.W., Weir, L.A., McClintock, B.T., Grant, E.H.C., Bailey, L.L. & Simons, T.R. (2012b) Experimental investigation of false positive errors in auditory species occurrence surveys. *Ecological Applications*, **22**, 1665–1674.
- Miller, D.A.W., Talley, B.L., Lips, K.R. & Grant, E.H.C. (2012c) Estimating patterns and drivers of infection prevalence and intensity when detection is imperfect and sampling errors occurs. *Methods in Ecology and Evolution*, **3**, 805–859.
- Miller, D.A., Nichols, J.D., Gude, J.A., Rich, L.N., Podrutzny, K.M., Hines, J.E. & Mitchell, M.S. (2013) Determining occurrence dynamics when false positives occur: estimating the range dynamics of wolves for public survey data. *PLoS ONE*, **8**, e65808 doi:10.1371/journal.pone.0065808.
- Molinari-Jobin, A., Kéry, M., Marboutin, E., Molinari, P., Koren, I., Fuxjäger, C. *et al.* (2012) Monitoring in the presence of species misidentification: the case of the Eurasian lynx in the Alps. *Animal Conservation*, **15**, 266–273.
- Mordecia, R.S., Mattsson, B.J., Tzilkowski, C.J. & Cooper, R.J. (2011) Addressing challenges when studying mobile or episodic species: hierarchical Bayes estimation of occupancy and use. *Journal of Applied Ecology*, **48**, 56–66.
- Murray, K.A., Skerratt, L.F., Spear, R. & McCallum, H. (2009) Impact and dynamics of disease in species threatened by the amphibian chytrid fungus, *Batrachochytrium dendrobatidis*. *Conservation Biology*, **23**, 1242–1252.
- Muths, E., Pedersen, B.S. & Pedersen, F.S. (2009) How relevant is opportunistic Bd sampling: are we ready for the big picture? *Herpetological Review*, **40**, 183–184.
- Muths, E., Corn, P.S., Pessier, A.P. & Green, D.E. (2003) Evidence for disease-related amphibian decline in Colorado. *Biological Conservation*, **110**, 357–365.
- Nichols, J.D., Hines, J.E., MacKenzie, D.I., Seamans, M.E. & Guitierrez, R.J. (2007) Occupancy estimation and modeling with multiple states and state uncertainty. *Ecology*, **88**, 1395–1400.
- 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 modeling using multiple detection methods. *Journal of Applied Ecology*, **45**, 1321–1329.
- Pavlacki, D.C., Blakesley, J.A., White, G.C., Hanni, D.J. & Lukacs, P.M. (2012) Hierarchical multi-scale occupancy estimation from monitoring wildlife populations. *Journal of Wildlife Management*, **76**, 154–162.
- Pilliod, D.S., Muths, E., Scherer, R.D., Bartelt, P.E., Corn, P.S., Hossack, B.R., Lambert, B.A., McCaffery, R. & Gaughan, C. (2010) Effects of amphibian chytrid fungus on individual survival probability in wild boreal toads. *Conservation Biology*, **24**, 1259–1267.
- Piotrowski, J.S., Annis, S.L. & Longcore, J.E. (2004) Physiology of *Batrachochytrium dendrobatidis*, a chytrid pathogen of amphibians. *Mycologia*, **96**, 9–15.
- Pollock, K.H. (1982) A capture-recapture design robust to unequal probability of capture. *Journal of Wildlife Management*, **46**, 752–757.

- Pradel, R. (2005) Multievent: an extension of multi-state capture-recapture models to uncertain states. *Biometrics*, **61**, 442–447.
- Rota, C.T., Fletcher, R.J., Dorazio, R.M. & Betts, M.G. (2009) Occupancy estimation and the closure assumption. *Journal of Applied Ecology*, **46**, 1173–1181.
- Royle, J.A. (2004) Modeling abundance index data from anuran calling surveys. *Conservation Biology*, **18**, 1378–1385.
- Royle, J.A. (2006) Site occupancy models with heterogeneous detection probabilities. *Biometrics*, **62**, 97–102.
- Royle, J.A. & Dorazio, R.M. (2008) *Hierarchical Modeling and Inference in Ecology*. Academic Press, San Diego, California.
- Royle, J.A. & Link, W.A. (2005) A general class of multinomial mixture models for anuran calling survey data. *Ecology*, **86**, 2505–2512.
- 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. & Nichols, J.D. (2003) Estimating abundance from repeated presence-absence data or point counts. *Ecology*, **84**, 777–790.
- Royle, J.A., Nichols, J.D. & Kéry, M. (2005) Modelling occurrence and abundance of species when detection is imperfect. *Oikos*, **110**, 353–359.
- Shea, C.P., Peterson, J.T., Wisniewski, J.M. & Johnson, N.A. (2011) Misidentification of freshwater mussel species (Bivalvia: Unionidae): contributing factors, management implications, and potential solutions. *Journal of the North American Benthological Society*, **30**, 446–458.
- Simons, T.R., Alldredge, M.W., Pollock, K.H. & Wettroth, J.M. (2007) Experimental analysis of the auditory detection process on avian point counts. *Auk*, **124**, 986–999.
- Voyles, J., Young, S., Berger, L., Campbell, C., Voyles, W.F., Dinudom, A. *et al.* (2009) Pathogenesis of Chytridiomycosis, a cause of catastrophic amphibian declines. *Science*, **326**, 582–585.
- Waddle, H., Thigpen, T.F. & Glorioso, B.M. (2009) Efficacy of automatic vocalization recognition software for anuran monitoring. *Herpetological Conservation and Biology*, **4**, 384–388.
- White, G.C. & Burnham, K.P. (1999) Program MARK: survival estimation from populations of marked animals. *Bird Study*, **46**, 120–139.
- Yackulic, C.B., Reid, J., Davis, R., Hines, J.E., Nichols, J.D. & Forsman, E. (2012) Neighborhood and habitat effects on vital rates: expansion of the Barred Owl in the Oregon Coast Ranges. *Ecology*, **93**, 1953–1966.
- Yackulic, C.B., Reid, J., Nichols, J.D., Hines, J.E., Davis, R. & Forsman, E. (in press) The roles of competition and habitat in the dynamics of populations and species distributions. *Ecology*. URL: <http://dx.doi.org/10.1890/13-0012.1>
- Zeller, K.A., Nijhawan, S., Salom-Perez, R., Potosme, S.H. & Hines, J.E. (2011) Integrating occupancy modeling and interview data for corridor identification: a case study for jaguars in Nicaragua. *Biological Conservation*, **144**, 892–901.
- Zipkin, E.F., Dewan, A. & Royle, J.A. (2009) Impact of forest fragmentation on species richness: a hierarchical approach to community modeling. *Journal of Applied Ecology*, **46**, 815–822.

Received 23 May 2013; accepted 18 July 2013

Handling Editor: Evan Cooch