ESTIMATING SITE OCCUPANCY, COLONIZATION, AND LOCAL EXTINCTION WHEN A SPECIES IS DETECTED IMPERFECTLY

Darryl I. MacKenzie,[1,5] James D. Nichols,[2] James E. Hines,[2] Melinda G. Knutson,[3]
and Alan B. Franklin[4]

[1]*Proteus Research and Consulting Ltd., P.O. Box 5193, Dunedin, New Zealand*
[2]*USGS, Patuxent Wildlife Research Center, Laurel, Maryland 20708-4017 USA*
[3]*USGS, Upper Midwest Environmental Sciences Center, La Crosse, Wisconsin 54603 USA*
[4]*Colorado Cooperative Fish and Wildlife Research Unit, Colorado State University, Fort Collins, Colorado 80523 USA*

Abstract. Few species are likely to be so evident that they will always be detected when present. Failing to allow for the possibility that a target species was present, but undetected, at a site will lead to biased estimates of site occupancy, colonization, and local extinction probabilities. These population vital rates are often of interest in long-term monitoring programs and metapopulation studies. We present a model that enables direct estimation of these parameters when the probability of detecting the species is less than 1. The model does not require any assumptions of process stationarity, as do some previous methods, but does require detection/nondetection data to be collected in a manner similar to Pollock's robust design as used in mark–recapture studies. Via simulation, we show that the model provides good estimates of parameters for most scenarios considered. We illustrate the method with data from monitoring programs of Northern Spotted Owls (*Strix occidentalis caurina*) in northern California and tiger salamanders (*Ambystoma tigrinum*) in Minnesota, USA.

Key words: colonization; detection probability; local extinction; metapopulation; monitoring; open population; patch occupancy; robust design; site occupancy.

INTRODUCTION

Estimating the proportion of sites occupied by a target species is important in both long-term monitoring programs and metapopulation studies. In a monitoring context, site occupancy probabilities may be used as a metric reflecting the current state of the population. Although estimates of abundance traditionally are used as a measure of system state (e.g., Williams et al. 2002), abundance estimation often requires more expense and effort than estimation of site occupancy (e.g., Tyre et al. 2001, MacKenzie et al. 2002). In metapopulation and island biogeographic studies, site (or patch) occupancy often is used as a state variable and incorporated into ''incidence functions'' that may be used to estimate local extinction and colonization probabilities (e.g., Hanski 1992, 1994, 1997, Moilanen 1999). In both contexts, an important, commonly overlooked fact is that the species often will not be detected even when present at a site. Although detection indicates that the species is present, nondetection does not equate to species absence. Failing to account for imperfect detectability will result in underestimates of site occupancy and biased estimates of local colonization and extinction probabilities. The species turnover rate will also be biased, as an apparent recolonization of a site

actually may be due to nondetection of the species at a previous sampling period.

MacKenzie et al. (2002) describe a method for estimating the proportion of sites occupied by a single species during a short time interval (e.g., a single breeding season) when the probability of detecting the species is less than 1. They envisage a sampling scheme where *N* sites are visited on each of *T* sampling occasions, and detection/nondetection of the species is recorded on each visit. A model based upon the detection histories for each site is constructed using straightforward probabilistic arguments and permits maximum likelihood estimation of the model parameters. MacKenzie et al. (2002) show that the sampling scheme can be relaxed somewhat to allow for missing values by a slight model adjustment, and that covariates such as habitat type, patch size, or other environmental variables can be easily incorporated. An important assumption of their model is that sites are closed to changes of the occupancy state during the sampling interval. This may be reasonable over a relatively short time interval (e.g., within a single year), but is unlikely to hold for longer studies (i.e., across multiple years). Indeed, in many situations the rate of change in site occupancies may be of more interest than the overall proportion of sites occupied at any given time. Here our interest is in long-term studies such as those using site occupancy surveys for biological monitoring or focusing on mechanisms underlying site occupancy dynamics.

The vital rates of site occupancy dynamics are local extinction and colonization probabilities. A popular approach to estimating these parameters has been based on so-called ''incidence functions,'' with incidence viewed as either the probability that a species occupies a specific site or the expected fraction of similar sites that are occupied. The pattern of apparently occupied sites at a single point in time is sometimes assumed to represent the equilibrium of a stationary Markov process, and local extinction and colonization probabilities are estimated based on this assumption (e.g., Hanski 1992, 1994, 1997). However, it is always difficult to infer process (extinction and colonization dynamics) from pattern (site occupancy); typically, many processes can produce any pattern. Clinchy et al. (2002) recently used computer simulation to demonstrate that observed patterns of patch occupancy by pikas (*Ochotona princeps*) could have been produced by very different processes and, hence, very different patterns of colonization and extinction.

Previous studies that have estimated local colonization and extinction probabilities using long-term data on site occupancy over many sampling periods (e.g., years) most commonly have assumed that detection probability is 1 (occupancy status is known with certainty in each sampling period) and that occupancy dynamics are generated by a stationary Markov process (Diamond and May 1977, Clark and Rosenzweig 1994, Rosenzweig and Clark 1994). Other recent approaches permit time-specific vital rates and thus make no assumption about process stationarity, but still assume perfect detection (Erwin et al. 1998, Moilanen 1999). We stress that neither assumption is required for our approach and believe such assumptions are unlikely to hold for most organisms.

Recently, Moilanen (2002) investigated the effects of various data error types on parameter estimates for Stochastic Patch Occupancy Models (SPOMs); the error with the most serious implications was the recording of false absences (because species were not detected with certainty). Moilenan (2002) suggested modifying the probabilities of transition between patch states (occupied and unoccupied) in successive years by incorporating a term into the SPOMs for the probability of observing a false absence. Yet Moilanen asserted that the parameter cannot be estimated from the type of data being considered, where each patch was surveyed for species presence/absence only once per year; ancillary information on detection probabilities is required. We note that this modification only partially allows for the effect of species nondetections. The connectivity measure used by Moilanen (2002) and many others, ($S_i(t)$), is a function of the observed state of the patches, which is not exactly known. To fully correct for false absences, the connectivity measure should also be modified, but that is not the scope of this paper.

We begin by outlining a basic sampling scheme that allows estimation of site occupancy, colonization, and local extinction probabilities when a species is not detected with certainty. We extend the methods of MacKenzie et al. (2002) and present a statistical model that permits direct estimation of these parameters. Simulation is used to assess model performance. To illustrate its utility, we analyze monitoring data for Northern Spotted Owls (*Strix occidentalis caurina*) in northern California and tiger salamanders (*Ambystoma tigrinum*) in Minnesota farm ponds. The freely downloadable program PRESENCE was used to analyze the data.[6]

METHODS

Basic sampling situation

We envisage a situation where N sites are surveyed over time with the intent to establish the presence or absence of a species. Sites may be a naturally occurring sampling unit such as a discrete pond or patch of vegetation, an investigator-defined monitoring station, or a quadrat chosen from a predefined area of interest. The occupancy state of sites may change over time, but there are intervals within the study during which it is reasonable to assume that, for all sites, no changes occur (e.g., annual breeding seasons in a 5-year study of migratory birds, or days within a week-long study of insects). The study consists of T primary sampling periods, between which changes in site occupancy state may occur. Within each primary period, investigators use an appropriate technique to detect the species at k_t surveys of the site. This is akin to Pollock's robust design for mark–recapture studies where k_t surveys represent secondary sampling periods within each primary period (Pollock 1982, Pollock et al. 1990).

The target species may or may not be detected at each survey and is not falsely detected when absent. The resulting detection history for each site may be expressed as T vectors of 1's and 0's, indicating detection and nondetection of the species, respectively. We denote the detection history for the k_t surveys of site i at primary sampling period t as $\mathbf{X}_{i,t}$, and the complete detection history for site i, over all primary periods, as $\mathbf{X}_i$. Detection histories for the N sites can be used to estimate site occupancy, colonization, and local extinction probabilities using a simple application of likelihood theory.

Statistical model

The situation presented by MacKenzie et al. (2002) could be considered as a special case of those considered here, where multiple surveys of the sites are only conducted for a single primary period. Let ψ_1 be the probability that a site is occupied by the species (at $t = 1$) and $p_{t,j}$ be the probability of detecting the species, given presence, in survey j within primary period t. MacKenzie et al. (2002) use a series of probabilistic arguments to model detection histories in the following

[6] URL: ⟨http://www.proteus.co.nz⟩

TABLE 1. Sample detection histories, $\mathbf{X}_i$, and the associated probabilities of observing them, $\Pr(\mathbf{X}_i)$.

$\mathbf{X}_i$	$\Pr(\mathbf{X}_i)$
101 110	$\phi_0 D(\mathbf{p}_{101,1})\phi_1\mathbf{p}_{110,2}$

$$= [\psi_1 \quad 1 - \psi_1]\begin{bmatrix} p_{1,1}(1 - p_{1,2})p_{1,3} & 0 \\ 0 & 0 \end{bmatrix}$$

$$\times \begin{bmatrix} 1 - \varepsilon_1 & \varepsilon_1 \\ \gamma_1 & 1 - \gamma_1 \end{bmatrix}\begin{bmatrix} p_{2,1}p_{2,2}(1 - p_{2,3}) \\ 0 \end{bmatrix}$$

$$= \psi_1 p_{1,1}(1 - p_{1,2})p_{1,3}(1 - \varepsilon_1)p_{2,1}p_{2,2}(1 - p_{2,3})$$

| 010 000 | $\phi_0 D(\mathbf{p}_{010,1})\phi_1\mathbf{p}_{000,2}$ |

$$= [\psi_1 \quad 1 - \psi_1]\begin{bmatrix} (1 - p_{1,1})p_{1,2}(1 - p_{1,3}) & 0 \\ 0 & 0 \end{bmatrix}$$

$$\times \begin{bmatrix} 1 - \varepsilon_1 & \varepsilon_1 \\ \gamma_1 & 1 - \gamma_1 \end{bmatrix}\begin{bmatrix} \prod_{j=1}^{3}(1 - p_{2,j}) \\ 1 \end{bmatrix}$$

$$= \psi_1(1 - p_{1,1})p_{1,2}(1 - p_{1,3})\left\{(1 - \varepsilon_1)\prod_{j=1}^{3}(1 - p_{2,j}) + \varepsilon_1\right\}$$

| 000 000 | $\phi_0 D(\mathbf{p}_{000,1})\phi_1\mathbf{p}_{000,2}$ |

$$= [\psi_1 \quad 1 - \psi_1]\begin{bmatrix} \prod_{j=1}^{3}(1 - p_{1,j}) & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 - \varepsilon_1 & \varepsilon_1 \\ \gamma_1 & 1 - \gamma_1 \end{bmatrix}$$

$$\times \begin{bmatrix} \prod_{j=1}^{3}(1 - p_{2,j}) \\ 1 \end{bmatrix}$$

$$= \psi_1 \prod_{j=1}^{3}(1 - p_{1,j})\left\{(1 - \varepsilon_1)\prod_{j=1}^{3}(1 - p_{2,j}) + \varepsilon_1\right\}$$

$$+ (1 - \psi_1)\left\{\gamma_1 \prod_{j=1}^{3}(1 - p_{2,j}) + (1 - \gamma_1)\right\}$$

Note: Detection histories represent the sequence of species detections (1) and nondetections (0), with groupings defining the primary surveying periods (e.g., years).

manner. If sites were surveyed three times to detect the species, the probability of obtaining the detection history 101 (i.e., the species was detected in surveys 1 and 3 but not in survey 2), can be modeled as follows:

$$\Pr(\mathbf{X}_{i,1} = 101) = \psi_1 p_{1,1}(1 - p_{1,2})p_{1,3}. \tag{1}$$

If the species was never detected at a site (history 000), there are two possibilities: either it was present but never detected, or it was genuinely absent. The probability of this history could be expressed as

$$\Pr(\mathbf{X}_{i,1} = 000) = \psi_1 \prod_{j=1}^{3}(1 - p_{1,j}) + (1 - \psi_1). \tag{2}$$

To extend this model beyond $T = 1$, we introduce two additional parameters. Let γ_t be the probability that a site unoccupied at primary sampling period t is occupied by the species at $t + 1$ (colonization). Let ε_t be the probability that a site occupied by the species at t is unoccupied at $t + 1$ (local extinction). With these

extra parameters we can continue to use a probabilistic argument for later primary sampling periods.

Consider the detection history 001 000: the species was detected at the site in the first primary period ($t = 1$) during the third survey only, and was never detected in the second primary period ($t = 2$). The probability of observing the detection history at $t = 1$ would be

$$\Pr(\mathbf{X}_{i,1} = 001) = \psi_1(1 - p_{i,1})(1 - p_{1,2})p_{1,3}. \tag{3}$$

Following this, there are two possibilities that would result in the species not being detected at $t = 2$: either it continued to occupy the site and was undetected, or it became locally extinct. The probability of this occurring would be expressed as

$$\Pr(\mathbf{X}_{i,2} = 000 \mid \mathbf{X}_{i,1}) = (1 - \varepsilon_1)\prod_{j=1}^{3}(1 - p_{2,j}) + \varepsilon_1. \tag{4}$$

The probability of observing the full detection history, $\Pr(\mathbf{X}_t = 001\ 000)$, would be the product of Eqs. 3 and 4.

Potentially, a number of different pathways could result in an observed detection history. It is thus most convenient to describe the model using matrix notation. Let ϕ_t be a 2×2 matrix of transition probabilities between states of occupancy from t to $t + 1$ at primary sampling periods $1, \ldots T - 1$, where rows of ϕ_t represent the occupancy state of the site at t (state 1, occupied; state 2, unoccupied) and columns represent the occupancy state at $t + 1$. Then,

$$\phi_i = \begin{bmatrix} 1 - \varepsilon_t & \varepsilon_t \\ \gamma_t & 1 - \gamma_t \end{bmatrix}.$$

Further, let ϕ_0 be the row vector

$$\phi_0 = [\psi_1 \quad 1 - \psi_1]$$

where ψ_1 is the probability that the site is occupied in the first primary sampling period. Let $\mathbf{p}_{X,t}$ be a column vector where each entry denotes the probability of observing the detection history $\mathbf{X}_{i,t}$ in primary period t, conditional upon occupancy state. For instance,

$$\mathbf{p}_{101,t} = \begin{bmatrix} p_{t,1}(1 - p_{t,2})p_{t,3} \\ 0 \end{bmatrix}$$

$$\mathbf{p}_{000,t} = \begin{bmatrix} \prod_{j=1}^{3}(1 - p_{t,j}) \\ 1 \end{bmatrix}.$$

Whenever the species is detected at least once during a primary period, the second element of $\mathbf{p}_{X,t}$ will always be zero, as it is impossible to observe such a history if the site is in the unoccupied state.

The probability for an observed detection history could then be calculated as

$$\Pr(\mathbf{X}_i) = \phi_0 \prod_{t=1}^{T-1} D(\mathbf{p}_{X,t})\phi_t\mathbf{p}_{X,T} \tag{5}$$

where $D(\mathbf{p}_{X,t})$ is a diagonal matrix with the elements of

$\mathbf{p}_{X,t}$ along the main diagonal (top left to bottom right), zero otherwise. Examples of observed detection histories and their probability statements, according to this model, are given in Table 1.

The model likelihood is then calculated according to Eq. 6, and the parameter values that maximize the likelihood can be obtained:

$$L(\psi_1, \boldsymbol{\varepsilon}, \boldsymbol{\gamma}, \mathbf{p} \mid \mathbf{X}_1, \ldots, \mathbf{X}_N) = \prod_{i=1}^{N} \Pr(\mathbf{X}_i). \qquad (6)$$

An important assumption of the model is that all parameters are constant across sites at any given time. Failure of this assumption introduces heterogeneity, which will lead to biased parameter estimates. Use of covariates enables this assumption to be relaxed somewhat, but heterogeneity created by unknown or unmeasured factors may be problematic.

Note that the probability of occupancy at time t, ψ_t, can be calculated recursively by the relationship

$$\psi_t = \psi_{t-1}(1 - \varepsilon_{t-1}) + (1 - \psi_{t-1})\gamma_{t-1}. \qquad (7)$$

Using this relationship, we can reparameterize the model to estimate ψ_t directly, by rearranging Eq. 7 to make either ε_{t-1} or γ_{t-1} the subject. Similarly, we imagine that the rate of change in occupancy may be of interest in some situations (e.g., change in the size of the species' range). By means of analogy with population size, we define such a rate of change as

$$\lambda_t = \frac{\psi_{t+1}}{\psi_t} \qquad (8)$$

for $t = 1, 2, \ldots, T - 1$. By rearranging Eq. 8 and substituting into Eq. 7, we can directly estimate the rate of change in site occupancy.

Finally, we note that our model is closely related to the modeling of temporary emigration in the mark–recapture context (Kendall et al. 1997). Indeed, such mark–recapture models recently were used to model colony site dynamics for colonial waterbirds (Barbraud et al. 2003). Although such an approach is conditional upon the species being detected at a site at least once and does not permit direct estimates of site occupancy probabilities.

Extensions to the model

Missing observations.—In many situations, there may be occasions when some sites cannot be surveyed for various reasons: a technician's vehicle may break down enroute to the site; logistic constraints may limit the number of sites that can be surveyed during a sampling period; or the sampling protocol may not require a site to be revisited in the primary period once the species has been detected there. In such instances, missing observations contribute no information about model parameters. This can be easily accounted for in the previous model.

If missing observations occur in the secondary sampling periods, the approach used by MacKenzie et al. (2002) can be applied: and the vector $\mathbf{p}_{X,t}$ is adjusted by removing the corresponding $p_{t,j}$ parameter(s). For example, if the history 11_ is obtained at primary period t (where "_" indicates a missing observation), then,

$$\mathbf{p}_{11\_,t} = \begin{bmatrix} p_{t,1}p_{t,2} \\ 0 \end{bmatrix}.$$

This represents that fact that no information, either detection or nondetection, was gathered at the third sampling occasion of primary period t.

Similarly, the model can be adjusted for situations in which a site was not surveyed for an entire primary period. Consider the following history, where the site was not surveyed in the second primary period: 101___001. Again, no information has been collected regarding detection or nondetection of the species. In this case however, we must allow for the fact that the occupancy state of the site is also completely unknown. This can be achieved by omitting $\mathbf{p}_{X,z}$ entirely; i.e., the probability of this detection history is

$$\Pr(\mathbf{X}_i = 101 \_\_\_ 001) = \boldsymbol{\phi}_0 D(\mathbf{p}_{101,1})\boldsymbol{\phi}_1\boldsymbol{\phi}_2\mathbf{p}_{001,3}. \qquad (9)$$

Including covariates.—Potential relationships between covariates and parameters can be investigated using the logistic model of Eq. 10, where θ is the probability of interest, $\mathbf{Y}$ is the matrix of covariate information, and $\boldsymbol{\beta}$ is the vector of logistic model coefficients to be estimated:

$$\theta = \frac{\exp(\mathbf{Y}\boldsymbol{\beta})}{1 + \exp(\mathbf{Y}\boldsymbol{\beta})}. \qquad (10)$$

Suitable covariates for site occupancy, colonization, and local extinction probabilities would be site-specific variables that may change with each primary sampling period (e.g., habitat type or generalized weather patterns such as drought or El Niño years). Such covariates could also be used for detection probabilities, and covariates that may change with each secondary sampling period (e.g., precipitation or air temperature) could also be considered. This is unlike the situation in mark–recapture, where time-varying, individual (site-specific) covariates cannot be used in the modeling of capture histories from open populations because the covariate value is unknown for occasions when the individual is not encountered. In the site occupancy setting, however, the covariate may be measured regardless of whether the species is detected. As mentioned previously, including covariates may account for some forms of heterogeneity.

Model comparisons

Using the general framework just described, one could develop a suite of potential models that may be reasonable explanations of observed data, i.e., time-

TABLE 2. Summary of model selection procedure and parameter estimates for the Northern Spotted Owl (*Strix occidentalis caurina*).

Model	ΔAIC	w	K	ψ_{1997}	$\hat{\gamma}_{1997}$	$\hat{\gamma}_{1998}$	$\hat{\gamma}_{1999}$	$\hat{\gamma}_{2000}$
$\psi(1997)\gamma(\text{year})\varepsilon(\cdot)p(\text{year},\cdot)$	0.00	0.57	11	0.63	0.11	0.07	0.39	0.09
$\psi(1997)\gamma(\cdot)\varepsilon(\cdot)p(\text{year},\cdot)$	1.43	0.28	8	0.62	0.18	0.18	0.18	0.18
$\psi(1997)\gamma(\text{year})\varepsilon(\text{year})p(\text{year},\cdot)$	3.54	0.10	14	0.63	0.11	0.07	0.39	0.12
$\psi(1997)\gamma(\cdot)\varepsilon(\text{year})p(\text{year},\cdot)$	4.92	0.05	11	0.62	0.18	0.18	0.18	0.18
$\psi(1997)\gamma(\text{year})\varepsilon(\cdot)p(\cdot,\cdot)$	9.81	0.00	7	0.64	0.10	0.07	0.37	0.15
$\psi(1997)\gamma(\cdot)\varepsilon(\cdot)p(\cdot,\cdot)$	11.22	0.00	4	0.63	0.18	0.18	0.18	0.18
$\psi(1997)\gamma(\text{year})\varepsilon(\text{year})p(\cdot,\cdot)$	12.15	0.00	10	0.64	0.10	0.06	0.37	0.16
$\psi(1997)\gamma(\cdot)\varepsilon(\text{year})p(\cdot,\cdot)$	13.86	0.00	7	0.63	0.18	0.18	0.18	0.18
Model-averaged estimates				0.63	0.13	0.11	0.32	0.12
(1 SE)				(0.066)	(0.071)	(0.077)	(0.134)	(0.084)
Naïve estimates				0.60	0.14	0.13	0.38	0.17

Notes: ΔAIC is the difference in AIC values between each model and the low-AIC model; w is the AIC model weight; and K is the number of parameters in the model. Model-averaged estimates of detection probability are $\{\hat{p}_{1997}, \hat{p}_{1998}, \hat{p}_{1999}, \hat{p}_{2000}, \hat{p}_{2001}\} = \{0.59, 0.52, 0.41, 0.38, 0.54\}$.

specific vs. time-constant extinction probabilities, or models with and without a covariate of interest. Models in the candidate set could be formulated to reflect competing hypotheses about the system under study. Selecting the better model(s) from the candidate set may proceed either by hypothesis testing (such as likelihood ratio tests) or information-theoretic methods (such as Akaike's Information Criterion, AIC).

Although our approach does not require the assumption of stationarity, one could test whether such an assumption would be reasonable by including a model in the candidate set that represents this situation. Using alternative parameterizations of the model (Eqs. 7 and 8), the process would be stationary when the occupancy probability is constant across time, or when $\lambda_t = \lambda = 1$. Such a model can be formally compared with a more general model permitting time-varying λ_t to evaluate the assumption of stationarity.

SIMULATION STUDY

Simulation methods

We conducted a simulation study to investigate the model's performance under a variety of likely scenarios. We generated data from the model such that all parameters were constant with respect to time. For a given local extinction probability (ε), the colonization probability (γ) was chosen such that the expected level of occupancy remained equal to its initial level (ψ_1), although in some situations, when ψ_1 and ε are high, this was not possible; hence γ was set to 1. This type of situation would represent a population of sites that are in equilibrium with respect to overall occupancy, i.e., the number of local extinctions would approximately equal the number of colonizations.

We considered the effects of six factors: number of sites (N) = 20, 50, or 100; ψ_1 = 0.2, 0.5, or 0.8; ε = 0.1, 0.3, or 0.5; number of primary periods (T) = 3, 5, or 10; number of surveys per primary period (k_t) = 2 or 5; and probability of detecting the species (given presence) in a survey ($p_{t,j}$) = 0.2 or 0.5. In all, we

considered a total of 324 scenarios, each repeated 1000 times. Estimated parameter values and standard errors were recorded each time. Standard errors were obtained from the variance–covariance matrix, which was calculated by inverting the numerically approximated matrix of second partial derivatives of the likelihood.

Simulation results

In some instances (generally ≪5%) the matrix of second partial derivatives could not be inverted; this tended to occur when parameters were estimated to be very close to 0 or 1. On such occasions, we disregarded parameter estimates and variances.

Generally, the parameter estimates appear to be largely unbiased except when both k_t and $p_{t,j}$ are small, in which case site occupancy and colonization probabilities tend to be overestimated when ψ_1 is low to moderate, and underestimated otherwise. Local extinction probabilities also tend to be overestimated. Increasing N, T, k_t, or $p_{t,j}$ improves both the accuracy and precision of estimated parameter values.

Estimated standard errors, obtained from the matrix of second partial derivatives, are in good agreement with nominal standard errors calculated from the repeated parameter estimates for each scenario. Full results of the simulations are presented in the Appendix.

We have not assessed the robustness of our model to violations of assumptions such as independent detection histories or heterogeneity in detection probabilities across sites. Undoubtedly violations of such assumptions will influence parameter estimates, but the severity of the resulting problems is unknown and a current area of research.

EXAMPLES

Northern spotted owls.—Potential spotted owl territories in northern California have been monitored for the presence of breeding pairs since 1985. Here we restrict our analysis to a subset of 55 sites surveyed annually between 1997 and 2001. Each site was sur-

TABLE 2. Extended.

$\hat{\varepsilon}_{1997}$	$\hat{\varepsilon}_{1998}$	$\hat{\varepsilon}_{1999}$	$\hat{\varepsilon}_{2000}$
0.14	0.14	0.14	0.14
0.14	0.14	0.14	0.14
0.09	0.13	0.36	0.12
0.09	0.13	0.24	0.12
0.15	0.15	0.15	0.15
0.15	0.15	0.15	0.15
0.09	0.15	0.26	0.11
0.09	0.15	0.26	0.11
0.13	0.14	0.17	0.14
(0.038)	(0.037)	(0.062)	(0.038)
0.12	0.19	0.28	0.13

veyed up to eight times (average 5.3 times) during a breeding season to determine whether the territory was occupied by a breeding pair. Survey techniques were consistent from year to year and followed an established protocol (Franklin et al. 1996). However, due to logistical constraints, surveys within years were not conducted simultaneously across sites.

Our model was fit to the data, as were a number of nested submodels that constrained colonization, local extinction, or detection probabilities to be constant across years. For simplicity, we assume that detection probabilities were constant for all surveys within years, although we believe this to be unlikely in practice. Table 2 presents parameter estimates and results of a model selection procedure based upon AIC, an information-theoretic approach to model selection, with smaller values indicating a more parsimonious model (see Burnham and Anderson [2002] for a thorough discourse on model selection and inference using such techniques). Models are denoted using a system similar to that commonly used in mark–recapture, where terms in parentheses indicate what factors are present for each associated parameter (Lebreton et al. 1992), i.e., the model $\psi(1997)\gamma(\cdot)\varepsilon(year)p(year, \cdot)$ has occupancy estimated for 1997 (the first primary period); time-constant colonization probabilities; year-specific local extinction probabilities; and probability of detecting the species during a single survey of a site that is constant within years but varies between years.

Table 2 provides very strong evidence that detection probabilities should be modeled as year specific (AIC weights for $p(\cdot, \cdot)$ models <1%, indicating little support for these models given the observed data). The most parsimonious model (with respect to AIC) is $\psi(1997)\gamma(year)\varepsilon(\cdot)p(year, \cdot)$, suggesting that colonization probabilities are year specific and local extinction probabilities are constant. However, the second- and third-ranked models also have substantial AIC weight; hence other interpretations should be considered. Model averaging, used to account for model selection uncertainty (Burnham and Anderson 2002), suggests that colonization probability tends to be ~0.12 except after 1999, when it appears to be much higher

(0.32), whereas local extinction probabilities tend to be approximately constant (~0.14). Based on these values, from Eqs. 7 and 8 we calculate an average rate of change in occupancy of 0.98 ± 0.04 (±1 SE), suggesting a reasonably static average level of occupancy over 5 years. In this example, all of the dynamic parameters are estimated to be similar, but smaller, than their naïve counterparts. This implies that even with a reasonably high probability of detecting a breeding pair at a site during the survey (given presence, due to many repeat surveys per season), by not explicitly accounting for detectability we would have concluded a higher turnover rate in territory use. This inference is consistent with the simulation results of Moilanen (2002).

Tiger salamanders.—Forty farm ponds and natural wetlands in southeastern Minnesota were surveyed for amphibian activity during spring and summer of 2000 and 2001. Dip net surveys were conducted at each visit for larvae and metamorphs of various amphibian species. Here we focus on detection/nondetection data for tiger salamanders, restricting our analysis to data collected from May until the end of the first week in August (as tiger salamander larvae and metaporphs may not be present outside this time frame, which would violate our within-season closure assumption). The 40 ponds were visited 5.0 times on average during 2000, and 39 of these ponds were resampled during 2001 with 3.0 visits on average per pond. Tiger salamanders were detected at 8 ponds in 2000 and 11 in 2001. Naïve estimates of occupancy (Table 3) suggest an apparent increase of 41%. We use the alternative parameterization of the model by rearranging Eq. 7 to directly estimate the occupancy probability in 2001, along with the probability of local extinction between 2000 and 2001. For the sake of simplicity, in this example we ignore some aspects of the sampling design and assume that detection probabilities are constant within each year. Full details of the study are available.[7]

Candidate models were fit to the data and compared using AIC (Table 3 summarizes results of the model selection procedure). There is very strong evidence that detection probabilities should be modeled as year specific [the $p(\cdot, \cdot)$ models again have <1% of the AIC weights], with the probability of detecting tiger salamanders during a visit to an occupied pond being (~) 0.4 higher in 2001. The most parsimonious model, $\psi(\cdot)\varepsilon(\cdot)p(year, \cdot)$, suggests that the overall level of occupancy is constant with some local extinctions, and hence, a similar number of colonizations. The second-ranked model, $\psi(year)\varepsilon(\cdot)p(year, \cdot)$, also has a substantial model weight, implying that there may have been a ~10% increase in the proportion of ponds occupied. Model averaging suggests that the overall level of occupancy is very similar in the two years, with a nonzero local extinction probability. Clearly, by allow-

[7] URL: ⟨http://www.umesc.usgs.gov/terrestrial/amphibians/mknutson_5003869.html⟩

TABLE 3. Summary of the model selection procedure and parameter estimates (with 1 SE in parentheses) for the tiger salamander (*Ambystoma tigrinum*).

Model	ΔAIC	w	K	$\hat{\psi}_{2000}$	$\hat{\psi}_{2001}$	$\hat{\varepsilon}_{2000}$	$\hat{p}_{2000}$	$\hat{p}_{2001}$
$\psi(\cdot)\varepsilon(\cdot)p(\text{year}, \cdot)$	0.00	0.65	4	0.28 (0.070)	0.28 (0.070)	0.27 (0.152)	0.27 (0.068)	0.66 (0.079)
$\psi(\text{year})\varepsilon(\cdot)p(\text{year}, \cdot)$	1.80	0.26	5	0.26 (0.086)	0.29 (0.076)	0.24 (0.161)	0.29 (0.074)	0.66 (0.081)
$\psi(\cdot)p(\text{year}, \cdot)$	4.22	0.08	3	0.34 (0.079)	0.34 (0.079)	0.00 (0.0)†	0.22 (0.049)	0.57 (0.074)
$\psi(\text{year})\varepsilon(\cdot)p(\cdot, \cdot)$	10.18	0.00	4	0.22 (0.069)	0.34 (0.089)	0.15 (0.181)	0.45 (0.059)	0.45 (0.059)
$\psi(\cdot)\varepsilon(\cdot)p(\cdot, \cdot)$	10.48	0.00	3	0.27 (0.066)	0.27 (0.066)	0.34 (0.144)	0.46 (0.058)	0.46 (0.058)
$\psi(\cdot)p(\cdot, \cdot)$	17.78	0.00	2	0.35 (0.079)	0.35 (0.079)	0.00 (0.0)†	0.36 (0.046)	0.36 (0.046)
Model-averaged estimates				0.28 (0.078)	0.29 (0.074)	0.24 (0.164)	0.27 (0.071)	0.65 (0.084)
Naïve estimates				0.20	0.28	0.25		

Note: Definitions are as in Table 2. The proportion of sites occupied each year and the local extinction probability were directly estimated; hence, the colonization probability is not estimated.

† The parameter value was fixed; hence, it was known without error.

ing for the fact that probability of detection is <1, a considerably different conclusion is reached concerning the occupancy probabilities in the 2 years compared to the naïve estimates. This difference is largely accounted for by the fact that detection probabilities appear to be unequal in 2000 and 2001.

The third- and sixth-ranked models, with constant ψ and no ε parameter, represent the situation where no changes in the occupancy state of ponds occur between years. Colonization and local extinction probabilities are zero, and the proposed model reduces to the closed population model of MacKenzie et al. (2002).

DISCUSSION

Requiring the detection/nondetection data to be collected in accordance with Pollock's (1982) robust design may seem restrictive to some. However, repeated surveys are required in order to estimate detection probabilities, which enable direct estimation of the other parameters. If all sites are surveyed only once per primary period, then detection probability will be confounded with the parameters of interest, resulting in estimates that are biased to an unknown degree and unknown direction. Although we have presented the concept of repeated surveys as discrete visits to the sites, in practice they may constitute two or more independent surveys conducted during a single visit by single or multiple observers. Also, the repeated surveys may be restricted to a subsample of sites in order to collect sufficient information for estimating detection probabilities, which can then be applied to those sites only visited once. We believe there is a great deal of flexibility in how the repeated surveys of sites may be carried out.

However, careful attention must be devoted to effects of the proposed study design on likely methods of analysis. If observers' abilities to detect the target species differ substantially and observers only survey a single site, a form of heterogeneity may be introduced that would bias parameter estimates. Where possible, observers should be randomly allocated to different sites on each survey occasion to minimize such an effect, or should be suitably trained to have similar detection abilities. Also, some study designs may restrict the types of models that may be considered. For instance, if during a primary period a site is not revisited after the species is first detected, models involving survey-specific detection probabilities cannot be fit to the data and detection probabilities must be assumed constant within a primary period.

We foresee numerous applications of this modeling framework. Incidence function models frequently assume functional relationships between patch-specific extinction probabilities and patch characteristics (e.g., size and perhaps species abundance), and between patch-specific colonization rates and isolation (e.g., as a function of distance to other patches or to a source population; see Hanski 1994, 1997, Moilanen 2002). In contrast, our models treat these relationships as hypotheses to be tested, enabling an objective evaluation of the strength of the relationships for extinction and colonization probabilities, without restrictive assumptions about process stationarity and perfect detection. One may also formally compare various functional forms for extinction and colonization probabilities. For example, do the expressions for these quantities frequently used in metapopulation studies (e.g., Hanski 1992, 1994, 1997, Moilanen 1999, 2002) provide a better description of the observed data than simply assuming all patches have the same probabilities? Or, does some other covariate relationship result in a superior model? We believe that the λ parameterization (directly modeling rate of change in occupancy; Eq. 8) of these models may be especially useful for monitoring programs by providing a metric reflecting change

in system state, a use similar to that of an abundance-based λ. For invasive species or for species hypothesized to be especially sensitive to climate and habitat change, the interpretation of λ as reflecting changes in the proportion of area occupied (range expansion or contraction) may be useful. Estimates and comparisons of competing models (e.g., time-invariant vs. time-varying λ) can be used to assess assumptions about process stationarity that are required by many other analyses and estimators commonly used in metapopulation studies.

Finally, we note that in our experience, a common argument for not adopting a formal statistical approach to modeling wildlife data is that the required assumptions are unrealistic. Many believe that a naïve approach, such as using a count of the number of sites where the species was detected or the number of animals encountered, is more robust and requires fewer assumptions. The converse is actually true. For the count to be a valid population metric, naïve approaches virtually always require even more restrictive assumptions than a statistically rigorous approach. For example, the assumption of equal detection probabilities for a group of similar sites is much less restrictive than the assumption that detection probabilities all equal 1. The effect of violating assumptions is nearly always more profound on naïve approaches that do not account for underlying sampling processes that contributed to the observed data (e.g., Jolly and Dickson 1983, Pollock et al. 1990, Williams et al. 2002).

Literature Cited

Barbraud, C., J. D. Nichols, J. E. Hines, and H. Hafner. 2003. Estimating rates of local extinction and colonization in colonial species. Oikos **101**:113–126.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and inference—a practical information-theoretic approach. Second edition. Springer-Verlag, New York, New York, USA.

Clark, C. W., and M. L. Rosenzweig. 1994. Extinction and colonization processes: parameter estimates from sporadic surveys. American Naturalist **143**:583–596.

Clinchy, M., D. T. Haydon, and A. T. Smith. 2002. Pattern does not equal process: what does patch occupancy really tell us about metapopulation dynamics? American Naturalist **159**:351–362.

Diamond, J. M., and R. M. May. 1977. Species turnover rates on islands: dependence on census interval. Science **197**: 266–270.

Erwin, R. M., J. D. Nichols, T. B. Eyler, D. B. Stotts, and B. R. Truitt. 1998. Modeling colony site dynamics: a case study of Gull-billed Terns (*Sterna nilotica*) in coastal Virginia. Auk **115**:970–978.

Franklin, A. B., D. R. Anderson, E. D. Forsman, K. P. Burnham, and F. W. Wagner. 1996. Methods for collecting and analyzing demographic data on the Northern Spotted Owl. Studies in Avian Biology **17**:12–20.

Hanski, I. 1992. Inferences from ecological incidence functions. American Naturalist **139**:657–662.

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

Hanski, I. 1997. Metapopulation dynamics: from concepts and observations to predictive models. Pages 69–91 *in* I. A. Hanski and M. E. Gilpin, editors. Metapopulation biology: ecology, genetics, and evolution. Academic Press, New York, New York, USA.

Jolly, G. M., and J. M. Dickson. 1983. The problem of unequal catchability in mark–recapture estimation of small mammal populations. Canadian Journal of Zoology **61**: 922–927.

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

Lebreton, J. D., K. P. Burnham, J. Clobert, and D. R. Anderson. 1992. Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. Ecological Monographs **62**:67–118.

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

Moilanen, A. 1999. Patch occupancy models of metapopulation dynamics: efficient parameter estimation using implicit statistical inference. Ecology **80**:1031–1043.

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

Pollock, K. H. 1982. A capture–recapture design robust to unequal probability of capture. Journal of Wildlife Management **46**:757–760.

Pollock, K. H., J. D. Nichols, C. Brownie, and J. E. Hines. 1990. Statistical inference for capture–recapture experiments. Wildlife Monographs 107.

Rosenzweig, M. L., and C. W. Clark. 1994. Island extinction rates from regular censuses. Conservation Biology **8**:491–494.

Tyre, A. J., H. P. Possingham, and D. B. Lindenmayer. 2001. Inferring process from pattern: can territory occupancy provide information about life history parameters. Ecological Applications **11**:1722–1737.

Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. Analysis and management of animal populations. Academic Press, San Diego, California, USA.

APPENDIX

Full results of the simulation study assessing the accuracy of the estimated parameters are available in ESA's Electronic Data Archive: *Ecological Archives* E084-054-A1.