

SAMPLING DESIGN TRADE-OFFS IN OCCUPANCY STUDIES WITH IMPERFECT DETECTION: EXAMPLES AND SOFTWARE

LARISSA L. BAILEY,^{1,3} JAMES E. HINES,¹ JAMES D. NICHOLS,¹ AND DARRYL I. MACKENZIE²

¹U.S. Geological Survey, Patuxent Wildlife Research Center, 12100 Beech Forest Road, Laurel, Maryland 20708-4017 USA

²Proteus Wildlife Research Consultants, PO Box 5193, Dunedin, New Zealand

Abstract. Researchers have used occupancy, or probability of occupancy, as a response or state variable in a variety of studies (e.g., habitat modeling), and occupancy is increasingly favored by numerous state, federal, and international agencies engaged in monitoring programs. Recent advances in estimation methods have emphasized that reliable inferences can be made from these types of studies if detection and occupancy probabilities are simultaneously estimated. The need for temporal replication at sampled sites to estimate detection probability creates a trade-off between spatial replication (number of sample sites distributed within the area of interest/inference) and temporal replication (number of repeated surveys at each site). Here, we discuss a suite of questions commonly encountered during the design phase of occupancy studies, and we describe software (program GENPRES) developed to allow investigators to easily explore design trade-offs focused on particularities of their study system and sampling limitations. We illustrate the utility of program GENPRES using an amphibian example from Greater Yellowstone National Park, USA.

Key words: *Ambystoma tigrinum*; *detection probability*; *monitoring*; *program GENPRES*; *site occupancy*; *survey design*; *tiger salamanders*; *Yellowstone National Park*.

INTRODUCTION

Various fields of ecology use occupancy, or probability that a patch is occupied by a species, as a state variable to address hypotheses about habitat associations (e.g., Scott et al. 2002), species distribution (e.g., Fisher and Shaffer 1996, Van Buskirk 2005), and metapopulation dynamics (e.g., Hames et al. 2001, Barbraud et al. 2003, Martinez-Solano et al. 2003). Often, occupancy is the state variable of interest to wildlife managers assessing the impacts of management actions (Mazerolle et al. 2005), and it is commonly the focal variable in long-term monitoring programs (Manley et al. 2004). Recent papers have emphasized that reliable inferences from these types of studies require estimating occupancy from detection–nondetection (presence–absence) data in a manner that deals with detection probabilities <1 (Moilanen 2002, MacKenzie et al. 2002, 2003, 2005, Gu and Swihart 2004). A key feature of these new estimation methods is that they generally require both spatial and temporal replication. The need for temporal replication at sampled sites to estimate detection probability creates a trade-off between spatial replication (number of sample sites distributed within the area of interest/inference) and temporal replication (number of repeated surveys at each site).

MacKenzie and Royle (2005) presented the first investigation of these trade-offs, and their findings provide some needed guidance for efficient design of

occupancy studies (also see Field et al. 2005). MacKenzie and Royle (2005) investigated three possible designs for single-season studies: “standard design,” in which each site is surveyed the same number of times; “double-sampling design,” in which a subset of sites is surveyed multiple times and the remaining sites are sampled only once; and “removal design,” in which sites are surveyed multiple times until the target species is detected. The authors found that there was an optimal number of repeated surveys for each design, regardless of whether the objective was to: (1) achieve a desired level of precision for minimal total survey effort; or (2) minimize the variance of the occupancy estimator ($\text{var}(\hat{\psi})$) for a given total number of surveys (MacKenzie and Royle 2005). This optimal number of surveys depended on the study design and the true occupancy and detection probability for the target species, but was independent of the number of study sites. In general, the standard design outperformed the double-sampling design, and the removal design was also promising if detection probability was believed to be constant across surveys.

MacKenzie and Royle (2005) considered their results to be a useful starting point for practitioners planning occupancy studies, but cautioned that their results were based on a simple model, namely one that assumed that occupancy probability was similar across all sites and that detection probability was constant over both time and space (sites). Furthermore, their results were based on asymptotic (large sample) methods and the results may not hold for studies with small sample sizes. These authors, and others, recommend that study designs be evaluated on a case-by-case basis, tailoring the design to

Manuscript received 31 December 2005; revised 1 May 2006; accepted 5 May 2006. Corresponding Editor: J. Van Buskirk.

³ E-mail: lbailey@usgs.gov

specific scientific goals, the biology of the target species, and logistical or economical constraints.

In this paper, we describe software (program GENPRES) developed to allow investigators to explore design trade-offs focused on the specifics of their study systems and attendant sampling limitations. The program permits investigation of both single-season and multiple-season situations, with one or more groups of sites that differ in their sampling intensity, occupancy, or detection probabilities. Users supply information on the model type (single- or multiple-season), construct a relevant sampling design using a flexible general framework, and provide realistic parameter values, based on either guesses or, preferably, on analysis of pilot data. Evaluations can be performed via analytic-numeric approximations (based on expected values of sample data) or simulations with either predefined or user-defined occupancy models. Detection histories are generated and can be saved as input files for multi-model evaluation using other software programs such as program MARK (White and Burnham 1999) or PRESENCE (MacKenzie et al. 2006). We illustrate the use of program GENPRES using an amphibian example from the Greater Yellowstone National Park Network.

BACKGROUND: OCCUPANCY MODELS AND PARAMETER ESTIMATION

Estimating occupancy probability while adjusting for imperfect detection has been the focus of several recent papers (MacKenzie et al. 2002, Tyre et al. 2003, Wintle et al. 2004). The MacKenzie et al. (2002) approach has proved to be the most flexible, and both single-season and multi-season models (MacKenzie et al. 2003) are included in estimation software programs PRESENCE and MARK.

These methods assume that s sites are selected, using some type of probability-based sampling, from a population of S possible sites within the area of interest. The sampling units, or "sites," may be defined by the investigator (e.g., habitat quadrats) or may be naturally occurring, discrete patches, such as the ponds used in our amphibian example. Appropriate methods are used to survey the sites multiple times each season, perhaps for multiple seasons (e.g., years). Detection and nondetection information is recorded for each survey of a site. Nondetection may arise if either the target species does not occupy the site or the investigator does not detect the species at an occupied site. Within a given season, sites are assumed to be closed to changes in occupancy (i.e., sites are either always occupied or unoccupied by the species), but this assumption may be relaxed, provided that any changes occur completely at random (MacKenzie et al. 2006). Between seasons (e.g., $t \rightarrow t + 1$), changes in occupancy may occur due to processes such as colonization and local extinction. Additional assumptions that apply to both single- and multi-season models include: (1) detections occur independently at sites; (2) occupancy and detection probabilities are similar across

sites and time, except when differences can be modeled with covariates (e.g., habitat features); and (3) the target species is identified correctly.

Single-season model and parameter definitions

MacKenzie et al. (2002, 2005, 2006) defined a probability-based model that consisted of two kinds of parameters: ψ represents the probability that a site is occupied by the target species, and p_j is the probability of detecting the species at an occupied site during the j th independent survey of a site. Assuming that detection histories from all s sites are independent, maximum likelihood methods are used to estimate occupancy and detection probability. Additionally, it is possible to model either occupancy or detection probability as a function of measured covariates using the logistic equation

$$\theta_i = \frac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{1 + \exp(\mathbf{X}_i\boldsymbol{\beta})}$$

where θ_i represents the parameter of interest for site i , \mathbf{X}_i is the row vector of covariate information for site i , and $\boldsymbol{\beta}$ is the column vector of coefficients to be estimated. Occupancy probability may be modeled as a function of site-specific covariates that do not change during the season (e.g., habitat type), whereas detection probability may be modeled as a function of either site-specific or survey-specific covariates (e.g., weather conditions or observer). The same modeling procedure also can be used with a Bayesian philosophy to statistical inference and can be easily implemented using Markov chain Monte Carlo methods.

Multi-season model and parameter definitions

MacKenzie et al. (2003, 2006) extended the single-season model by introducing two vital rate parameters that govern changes in the occupancy state between successive seasons: ϵ_t represents the probability that an occupied site in season t becomes unoccupied in season $t + 1$ (i.e., the species goes locally extinct), and γ_t represents the probability that an unoccupied site in season t is occupied by the species in season $t + 1$ (colonization). The extinction and colonization processes are explicitly incorporated into a general model that also includes detection probability (MacKenzie et al. 2003, 2005a). The multi-season model is also likelihood based and parameters may be modeled as functions of measured covariates. Additionally, both single- and multiple-season models can accommodate "missing data," as illustrated in our subsequent examples in which some sites are visited less often than others. MacKenzie et al. (2005a) provide an excellent resource for more details on occupancy models.

PROGRAM GENPRES

Analytic-numeric approximations and simulations

Program GENPRES offers practitioners two methods for evaluating bias and precision of occupancy estima-

tors under different analysis models: simulations and analytic-numeric approximations. The analytic-numeric method involves using a known, “true” model with realistic parameter values to generate an artificial data set consisting of the expected value of each detection history under the assumed model (Nichols et al. 1981, Burnham et al. 1987:214–217, 292–295, Gimenez et al. 2004). Next, the generated data are analyzed as if they were real data, under any model of interest. The method is strictly numerical, but “analytical,” not Monte Carlo (Burnham et al. 1987:215). With large sample sizes, this method can be used to approximate estimator bias and precision, and power of likelihood ratio tests (Burnham et al. 1987:214–217, Gimenez et al. 2004). Program GENPRES aids the investigator in these tasks (data generation and analysis) for a wide variety of user-defined model and parameter combinations (examples will be discussed). The validity of the numeric-analytic approximations of bias and precision depends on the sample size being large (Nichols et al. 1981, Burnham et al. 1987:216); this method can give poor results if the design includes a small number of sites.

To examine small-sample properties, or to obtain empirical sampling distributions of estimators, users can select the “simulations” option and input the number of simulations they wish to run. Sequences of random numbers are compared with input parameter values in order to generate simulated detection history data. Resulting detection history data sets thus differ, despite being generated by the same input parameters and model. The model of interest is fit to each data set, and maximum likelihood estimates are thus obtained. The distribution of resulting estimates, and statistics computed from this distribution (e.g., standard deviation), permit evaluation of estimator performance, even in the case of small sample sizes.

Overview of program GENPRES

GENPRES is written in C and RAPIDQ (a visual BASIC variant) and can be downloaded from the Patuxent Wildlife Research Center website (*available online*).⁴ Input information required for program GENPRES includes the generating model type and structure, parameter values, analysis model(s), and evaluation method (either simulations or expected values; Fig. 1). The choice of the generating model type (single- or multi-season) defines the scope of parameters that are supplied by the user. Occupancy and detection parameters are required by both model types; if the user chooses a multi-season generating model, fields appear for the probability that an occupied site remains occupied ($\phi = 1 - \epsilon$) and the probability that an unoccupied site becomes colonized (γ).

Ecologists can investigate the impact of unmodeled parameter heterogeneity or the power of a given study

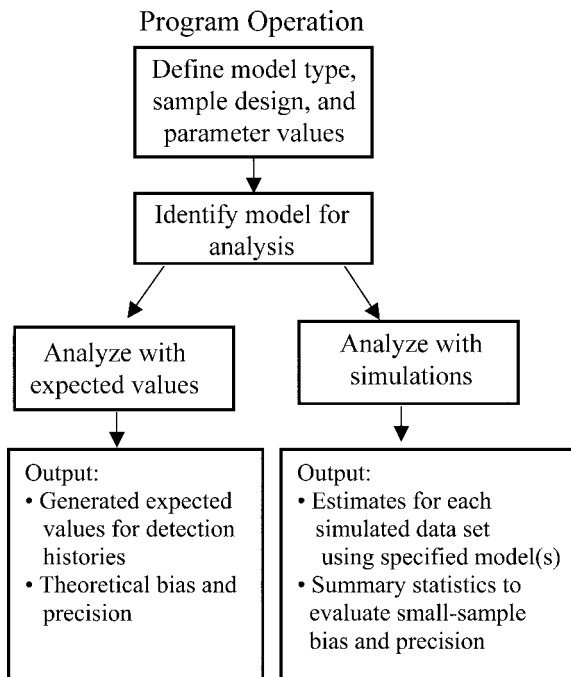


FIG. 1. Process used by practitioners to explore precision and bias of estimators in occupancy models.

design to detect parameter differences by including multiple groups of sites. Sites within each group are assumed to have the same parameter values, but these parameters may vary among groups. For example, an investigator may be interested in exploring variation of occupancy probabilities among two habitat types; here, each habitat type defines a group, and one might ask: what is the power to detect a 0.20 difference in occupancy probabilities between the two habitats for a fixed number of sites in each habitat (group)? Alternatively, resource managers can investigate trade-offs in estimator precision for different survey designs, where each group of sites may have a different survey frequency (examples follow). Program GENPRES requires the following information for each group of sites: the number of sites, the number of surveys, and relevant parameter values. Next, users identify models for analysis from a list of predefined models or construct their own models under a “user-defined” option. Finally, users may evaluate properties of estimators with analytic-numeric approximations (expected values) or simulations (Fig. 1). Output using the analytic-numeric approximation method can be used to approximate bias $[E(\hat{\theta}) - \theta]$, relative bias $[E(\hat{\theta}) - \theta]/\theta$, root mean square error $\sqrt{[E(\hat{\theta}) - \theta]^2}$, or similar metrics, where $\hat{\theta}$ denotes an estimator and θ the true parameter value. Simulation results, consisting of parameter estimates and standard errors calculated from each simulated data set, can be used to assess bias and precision of estimators using these metrics, except that $E(\hat{\theta})$ is estimated using the arithmetic mean of estimates

⁴ <http://www.mbr-pwrc.usgs.gov/software/>

from each simulation, $\bar{\hat{\theta}}$. The sampling variances (or standard deviations) of parameter estimates, $\text{var}(\hat{\theta})$, estimate the true sampling variances and can be compared to the arithmetic mean of the model-based asymptotic variance estimates, $\text{var}(\hat{\theta})$, to assess possible bias in the variance estimators. Mean squared error or root mean square error (RMSE) is often used to make comparisons among estimators because it represents the sum of variance and squared bias (Cochran 1977).

EXAMPLE: AMPHIBIANS IN GREATER
YELLOWSTONE NETWORK

Ecologists engaged in occupancy studies are often faced with issues of survey design, whether it be determining the number of cameras and their respective deployment time for studies of elusive mammals (MacKenzie et al. 2005, O'Connell et al. 2006), addressing potential observer bias for a cryptic insect species (MacKenzie et al. 2006:116–122), or developing multi-species avian monitoring programs that inform managers in time to implement management action (Field et al. 2005). Here we use one scenario from Yellowstone National Park, USA, to illustrate common questions that arise during the planning stages of large-scale studies or long-term monitoring programs. Our purpose is to demonstrate how program GENPRES can be used to address sampling design issues; therefore, although the scenario and pilot data for amphibians in Yellowstone National Park are realistic, we have not represented all of the complexities present in this system.

Pilot data description

Scientists associated with the USGS Amphibian Research and Monitoring Initiative (ARMI) and the National Park Service have gathered preliminary data on amphibian occurrence and distribution in various drainage catchments within Yellowstone National Park (Corn et al. 2005). Water bodies in one catchment have been surveyed during multiple seasons (years) from 2002 to 2004, using visual encounter surveys (for details on sampling protocol, see Corn et al. [2005] and Muths et al. [2005]). Between 44 and 77 water bodies (i.e., sites) were visited each season, but not all sites were resurveyed within a season and some sites were unavailable for amphibian occupancy during dry seasons. Species-specific detection data from each season were analyzed separately using single-season occupancy models (MacKenzie et al. 2002). The tiger salamander, *Ambystoma tigrinum*, consistently showed an increase in detection probabilities throughout the season, with estimates ranging from $\hat{p} = 0.12$ – 0.33 for early surveys in June to $\hat{p} \geq 0.70$ at the end of July. Occupancy probabilities for available (wet) sites ranged from ~ 0.35 to 0.75 for this species. Assuming that these data are representative of a larger area of interest, we use similar parameter values and patterns to address three common design questions. In each case, we are interested in: (1) comparing precision and bias of estimators under

different sampling designs and analysis models and (2) exploring whether sampling designs differ in their robustness to model misspecification. In addition to these shared objectives, the scenarios represent real studies with differing goals, hypotheses, and logistical limitations that may influence study design recommendations.

Single-season example: temporal trade-offs

Repeated surveys at sites can be accomplished in multiple ways; for example, a site may be surveyed by a single observer on different days, or a site could be surveyed by multiple independent observers on the same day (MacKenzie et al. 2002, 2006). In some situations, the modeling of data for these two different sampling designs would be identical, whereas in other situations the modeling might differ. For example, if multiple visits by the same investigator to a site could not be viewed as independent (the investigator retained knowledge of where to look for the species), then modeling would have to incorporate different detection probabilities for initial and subsequent detections. In situations where temporary emigration from the site is possible, then random emigration will cause parameter definitions to change slightly (Mackenzie et al. 2006:105–106).

In large, remote areas, such as Yellowstone National Park, observers usually work in groups for logistical and safety concerns; thus multiple independent observers are a logical choice for a survey method. However, there was concern among investigators that these data may not be sufficient to model occupancy in the presence of temporal changes in detection probability within each season. Although we were confident that temporal variation in detection probability would not translate into biased occupancy estimates if detection were properly modeled, we were interested in bias resulting from poor modeling of variation in detection probability. Thus, we examined three designs being considered by the investigators (Table 1) and assessed performance of estimators when detection probability was properly modeled and when it was not. Importantly, note that although the designs differ in the level of spatial and temporal replication, the total number of surveys is the same for all three designs. We assumed that enough resources are available for two observers to visit 12 sites every two weeks; a typical survey season in Yellowstone consists of four possible biweekly survey periods (eight possible surveys: four biweekly periods \times two observers). Using information obtained from the analysis of the pilot data, a reasonable occupancy probability was set at $\psi = 0.60$, and we anticipated negligible detection differences among observers. We assumed that detection probability could be expressed using the following logit function: $\text{logit}(p) = \beta_0 + \beta_1 \times (\text{biweekly survey period})$. Use of $\beta_0 \approx -1.80$ and $\beta_1 \approx 0.95$ yields $p_{1\&2} = 0.30$, $p_{3\&4} = 0.53$, $p_{5\&6} = 0.75$, $p_{7\&8} = 0.88$, which mimics estimates obtained from the pilot data. Here, p is detection probability, and the subscripts indicate sequential survey

TABLE 1. Representation of three designs used to explore trade-offs between spatial and temporal replication for a single-season occupancy example: distribution of sampling effort across four biweekly survey periods.

Design 1 ($s = 48$ sites)					Design 2 ($s = 30$ sites)					Design 3 ($s = 36$ sites)				
No. sites	Survey period				No. sites	Survey period				No. sites	Survey period			
	1	2	3	4		1	2	3	4		1	2	3	4
12	xx	—	—	—	6	xx	xx	xx	xx	6	xx	—	xx	—
12	—	xx	—	—	6	xx	—	—	—	6	—	xx	—	xx
12	—	—	xx	—	6	—	xx	—	—	6	xx	—	—	—
12	—	—	—	xx	6	—	—	xx	—	6	—	xx	—	—
					6	—	—	—	xx	6	—	—	xx	—
										6	—	—	—	xx

Notes: Each design assumes that two independent observers conduct surveys biweekly (every two weeks); each x denotes a survey by a single observer. Dashes indicate that those sites were not visited during those periods. Designs differ in the total number of sites and the temporal replication at each site, but the total number of surveys (98) is the same for all designs.

by two (numbered) observers. The three survey designs were constructed within GENPRES by grouping sites according to sampling frequency and detection probability.

First, we analyzed each design with large-sample numeric-analytic approximations (using expected values of detection history data). Table 2 contains occupancy estimates and standard errors for six candidate models that vary in their detection structure. Three models are consistent with the generating detection function: $\psi(\cdot)p(\text{linear})$ is the “true” generating model (i.e., $\text{logit}(p)$ is linear on a biweekly scale), $\psi(\cdot)p(\text{biweek})$ has different detection parameters for each two-week period, and $\psi(\cdot)p(t)$ has eight separate detection parameters. Notice that all three of these models have the same value for $-2\log(\text{likelihood})$. The remaining three models are inconsistent with the generating model: $\psi(\cdot)p(\cdot)$ denotes a model with constant detection probability across all surveys, while $\psi(\cdot)p(\text{month})$ and $\psi(\cdot)p(\text{obs})$ represent models with detection varying among months and observers, respectively.

Next, we used simulations to explore how well the large-sample numeric-analytic approximations performed for our study containing relatively small sample sizes ($s \leq 48$). Appendix A contains the command lines used to construct analysis models not found in the

predefined list. Simulation results are based on 1000 iterations for each design \times analysis model combination and are presented in Table 3.

Focusing on results from models consistent with the generating data, occupancy estimators are approximately unbiased for all three sampling designs, as expected (Table 2). Theoretical standard errors for the occupancy estimator are slightly better for designs 1 and 3 ($CV \approx 16\%$) compared to design 2 ($CV = 17.8\%$), but precision of detection probability estimates is better using designs 2 and 3 (results not shown). Design 1 appears less robust to model misspecification, producing greater estimator bias under models that are inconsistent with the generating data (Table 2).

These same general findings hold true in the simulation evaluations, but simulations also reveal some bias in estimators based on good approximating models (Table 3). Relative bias calculated for the occupancy estimator under model $\psi(\cdot)p(\text{linear})$ is +3.5% for design 1, +1.3% for design 2, and +2.6% for design 3 (Table 3). Using model $\psi(\cdot)p(\text{biweek})$, the relative bias in occupancy is: +6.0%, +4.3%, and +0.8% for designs 1, 2, and 3, respectively. All designs suggest that models that appropriately contain time variation in detection probability may slightly overestimate occupancy, whereas models lacking time-specific detection probabilities

TABLE 2. Summary of models fit to expected values of detection histories under three designs shown in Table 1.

Model	np	Design 1, 48 sites			Design 2, 30 sites			Design 3, 36 sites		
		$-2l$	$\hat{\psi}$	$\widehat{SE}(\hat{\psi})$	$-2l$	$\hat{\psi}$	$\widehat{SE}(\hat{\psi})$	$-2l$	$\hat{\psi}$	$\widehat{SE}(\hat{\psi})$
$\psi(\cdot)p(\text{linear})$	3	106.28	0.600	0.097	93.34	0.600	0.107	100.37	0.600	0.096
$\psi(\cdot)p(\text{month})$	3	107.94	0.589	0.097	95.38	0.595	0.108	102.30	0.594	0.096
$\psi(\cdot)p(\text{biweek})$	5	106.28	0.600	0.097	93.34	0.600	0.107	100.37	0.600	0.096
$\psi(\cdot)p(\cdot)$	2	113.46	0.532	0.086	103.48	0.566	0.105	110.44	0.573	0.096
$\psi(\cdot)p(\text{obs})$	3	113.46	0.532	0.086	103.48	0.566	0.105	110.44	0.573	0.096
$\psi(\cdot)p(t)$	9	106.28	0.600	0.097	93.34	0.600	0.107	100.37	0.600	0.096

Notes: Given for each model is the number of parameters (np), and under each design, twice the negative log-likelihood ($-2l$), estimates of occupancy ($\hat{\psi}$), and their corresponding standard errors. True probability of occupancy is 0.60. Models highlighted in boldface are consistent with the true generating model, $\psi(\cdot)p(\text{linear})$. Model $\psi(\cdot)p(\text{biweek})$ has different detection parameters for each two-week period; $\psi(\cdot)p(t)$ has eight separate detection parameters; $\psi(\cdot)p(\cdot)$ denotes a model with constant detection probability across all surveys; $\psi(\cdot)p(\text{month})$ and $\psi(\cdot)p(\text{obs})$ represent models with detection varying among months and observers, respectively.

TABLE 3. Simulation results for three sampling designs for specified analysis models (1000 replications each).

Model	np	Design 1, 48 sites				Design 2, 30 sites				Design 3, 36 sites			
		$\hat{\psi}$	$\widehat{SE}(\hat{\psi})$	$\overline{SE}(\hat{\psi})$	RMSE	$\hat{\psi}$	$\widehat{SE}(\hat{\psi})$	$\overline{SE}(\hat{\psi})$	RMSE	$\hat{\psi}$	$\widehat{SE}(\hat{\psi})$	$\overline{SE}(\hat{\psi})$	RMSE
$\psi(\cdot) p(\text{linear})$	3	0.621	0.101	0.097	0.103	0.608	0.104	0.105	0.105	0.616	0.096	0.096	0.097
$\psi(\cdot) p(\text{month})$	3	0.606	0.099	0.097	0.099	0.605	0.108	0.106	0.108	0.608	0.098	0.095	0.098
$\psi(\cdot) p(\text{biweek})$	5	0.636	0.102	0.101	0.108	0.626	0.113	0.105	0.116	0.605	0.055	0.056	0.055
$\psi(\cdot) p(\cdot)$	2	0.543	0.088	0.090	0.105	0.572	0.104	0.105	0.108	0.582	0.096	0.096	0.097
$\psi(\cdot) p(\text{obs})$	3	0.538	0.089	0.089	0.109	0.573	0.102	0.104	0.105	0.581	0.099	0.096	0.101
$\psi(\cdot) p(t)$	9	0.617	0.105	0.096	0.107	0.605	0.111	0.104	0.111	0.610	0.100	0.095	0.101

Notes: The generating model had occupancy probability $\psi = 0.60$ and detection probabilities that were equivalent among two observers but varied over four biweekly survey periods: $p = 0.30, 0.53, 0.75,$ and 0.88 . Reported estimates include: average occupancy ($\hat{\psi}$), estimated true sampling variance [reported as $\widehat{SE}(\hat{\psi})$, the estimated standard error of the occupancy estimates], the average of the asymptotic standard errors [denoted $\overline{SE}(\hat{\psi})$], and root mean square error (RMSE). Models highlighted in boldface are consistent with the true generating model; np is the number of parameters.

underestimate occupancy to a greater degree. Consistent with analytic-numeric methods, precision of the occupancy estimator is slightly better for design 1, but this is not necessarily encouraging because occupancy estimators have greater bias under design 1. RMSE, a metric that combines precision and bias, suggests that design 3 performs slightly better than either of the other two designs (Table 3). Together these findings suggest that sampling more sites with minimal temporal replication usually is not the best policy in planning occupancy studies; a finding echoed by MacKenzie and Royle (2005) and Field et al. (2005).

Multiple-season examples: allocation over space and time

Many occupancy studies and monitoring programs are planned for multiple seasons, where objectives focus on vital rates (colonization and extinction) and temporal and spatial factors that affect these rates. Using our tiger salamander example from Yellowstone National Park, we explored two general study design questions. The first question focuses again on how investigators might best allocate their survey effort over multiple seasons in order to maximize the precision of vital rate estimators. The second question focuses on how sites might be allocated among groups of sites with different habitat types or treatments that are believed to affect vital rates.

Focusing on our first question, we note that others have investigated this issue (MacKenzie 2005, MacKenzie et al. 2006). These authors used simulations to explore the relative benefits of using a “standard design” in which all sites are surveyed each season vs. a “rotating

panel design,” in which only a subset of sites is surveyed every season and the remaining sites are surveyed less frequently (e.g., surveyed every fifth season). Rotating panel designs are favored by some investigators because the spatial coverage of a study can be increased (i.e., the total number of sites surveyed at least once every five years is greater under the rotating panel design). However, others are critical of this type of design because increasing the spatial replication in this manner does not increase the effective sample size (e.g., for the purpose of estimating a trend in occupancy) and spatial and temporal effects may be confounded (MacKenzie et al. 2006).

Yellowstone National Park is included in a large amphibian study in the Rocky Mountain region (Corn et al. 2005). Investigators were trying to decide between two sampling designs: a standard design (design 1) in which the same 24 sites are surveyed every season vs. a rotating panel design (design 2) in which only 12 sites are surveyed every year, but 36 additional sites would be surveyed every other year (Table 4). Note that under design 2, the total number of surveys conducted is 25% greater than under design 1. In the rotating panel design, sampling effort would be concentrated in Yellowstone National Park one year and in other parks the following year. We explored the trade-offs in terms of precision and bias of vital rate estimators under these two designs. We generated these designs for a four-season tiger salamander study with the following parameter values: initial occupancy was $\psi = 0.60$ and time-constant extinction and colonization probabilities were $\epsilon = 0.25$

TABLE 4. Representation of two designs used to explore trade-offs between spatial and temporal replication for a multiple-season example: distribution of sampling effort across four seasons.

Design 1: Standard (24 sites, 192 surveys)					Design 2: Rotating panel (48 sites, 240 surveys)				
No. sites	Season				No. sites	Season			
	1	2	3	4		1	2	3	4
24	xx	xx	xx	xx	12	xx	xx	xx	xx
					36	—	xx	—	xx

Notes: Each x denotes a survey by a single independent observer. Dashes indicate that those sites were not visited during those periods.

TABLE 5. Summary of multi-season models fit to expected values of detection histories for two designs: standard and rotating panel design.

Model	np	Design 1, standard ($s = 24$ sites)							Design 2, rotating panel ($s = 48$ sites)						
		$-2l$	$\hat{\psi}$	$\widehat{SE}(\hat{\psi})$	$\hat{\epsilon}$	$\widehat{SE}(\hat{\epsilon})$	$\hat{\gamma}$	$\widehat{SE}(\hat{\gamma})$	$-2l$	$\hat{\psi}$	$\widehat{SE}(\hat{\psi})$	$\hat{\epsilon}$	$\widehat{SE}(\hat{\epsilon})$	$\hat{\gamma}$	$\widehat{SE}(\hat{\gamma})$
$\psi(\cdot) \epsilon(\cdot) \gamma(\cdot) p(j)$	5	184.29	0.600	0.135	0.250	0.114	0.200	0.097	233.95	0.600	0.146	0.250	0.106	0.200	0.091
$\psi(\cdot) \epsilon(t) \gamma(t) p(j)$	9	184.29	0.600	0.142	0.250†	0.175†	0.200†	0.159†	233.95	0.600	0.181	0.250†	0.199†	0.200†	0.180†
$\psi(\cdot) \epsilon(t) \gamma(t) p(tj)$	15	184.29	0.600	0.187	0.250†	0.198†	0.200†	0.183†	233.95	0.600	0.259	0.250†	0.213†	0.200†	0.199†
$\psi(\cdot) \epsilon(\cdot) \gamma(\cdot) p(t)$	7	208.39	0.703	0.159	0.155	0.107	0.143	0.132	262.67	0.723	0.193	0.129	0.110	0.135	0.164

Notes: The number of parameters (np) is given for each model. Also presented for each model under each design are twice the negative log-likelihood ($-2l$) and estimates of initial occupancy ($\hat{\psi}$), colonization ($\hat{\gamma}$), and extinction probabilities ($\hat{\epsilon}$), with their corresponding standard errors. True probabilities are: initial occupancy $\psi = 0.60$, time-constant vital rates $\epsilon_t = 0.25$ and $\gamma_t = 0.20$, and time-specific detection probabilities $p_1 = 0.25$ (early season) and $p_2 = 0.75$ (late season); j represents each independent survey of a site. Boldface models are consistent with the true generating model.

† The time-specific estimate exhibiting the smallest bias and greatest precision.

and $\gamma = 0.20$, respectively. These vital rate probabilities applied over four years produce a 22% decline from the initial occupancy level ($\psi_4 = 0.47$). For simplicity, we assume that designated sites were visited twice within a season by single, independent observers. Early-season detection probability was assumed to be $p_1 = 0.25$, and late-season detection probability was $p_2 = 0.75$.

Table 5 contains parameter and standard error estimates for four candidate models. Three of the four models are nested (parameters of the less general models can be obtained by constraining parameters of the more general), and the data were generated from the least general of these models. Thus, large-sample approximations of $-2 \log(\text{likelihood})$ are the same for these models. Of these three models, precision of estimators under a model assuming constant extinction and colonization is not expected to be worse under a rotating panel design, and this expectation is upheld (Table 5). However, in cases where extinction and colonization are modeled as time specific, the reduction in number of sites in years 1 and 3 under the rotating panel was expected to yield less precise estimates, and indeed this expectation was also shown to be reasonable (Table 5). The standard design also appears to produce estimators that are more robust to model misspecification, but both designs produce severe, negative bias in vital rate estimates if variation in detection probability is not included in the analysis model.

The generated detection histories and expected values may be analyzed under other parameterizations available in program MARK to allow season-specific occupancy estimates (results not shown). When we applied those models to these two sampling designs, our findings agreed with results from MacKenzie (2005): namely, that the key determinant of the precision of the season-specific occupancy estimate was the number of sites surveyed within the season, not the total number of sites surveyed over the duration of the study. Thus, in seasons 2 and 4, when 48 sites were sampled under design 2, occupancy estimates were more precise than under design 1 with only 24 sites surveyed. However, the opposite is true of occupancy estimators in seasons 1 and 3, when more sites were surveyed under design 1.

Another common objective in multi-season occupancy studies is to test a priori hypotheses about factors that may affect the vital rates that are responsible for population change. For example, in Yellowstone National Park, biologists have observed some amphibian species more frequently at sites influenced by geothermal activity (Koch and Peterson 1995). These sites generally have higher pH, conductivity, and acid-neutralizing capacity than other sites, perhaps allowing some resistance to acidification or disease (Koch and Peterson 1995). Alternatively, these sites, and the terrestrial habitat around them, may serve as refuges during severe winters. The following exercise may be performed by investigators interested in sample-size requirements to test hypotheses about differences in vital rates among sites with and without known geothermic influence. Suppose biologists believe that occupancy probabilities of tiger salamanders are fairly high on geothermally influenced sites (GS) and that these populations are quite stable, with low extinction probabilities. Parameter values for GS sites might be set at: initial occupancy $\psi_{GS} = 0.70$ and extinction probability $\epsilon_{GS} = 0.10$ ($\phi_{GS} = 1 - \epsilon_{GS} = 0.90$). Assuming that the system is near equilibrium, colonization probability could be calculated using the recursive equation, $\psi_{t+1} = \psi_t(1 - \epsilon_t) + (1 - \psi_t)\gamma_t$, yielding $\gamma_{GS} = 0.23$. Notice that under these parameter values, occupancy levels remain constant across all four seasons. Suppose pilot data are available suggesting that sites without geothermal influence (NGS) have lower occupancy probabilities for tiger salamanders (e.g., $\psi_{NGS} = 0.50$), and it is believed that extinction probabilities at these sites may be three times higher than on GS sites. Then parameter values for NGS sites may be: initial occupancy $\psi_{NGS} = 0.50$, extinction probability $\epsilon_{NGS} = 0.30$ ($\phi_{NGS} = 1 - \epsilon_{NGS} = 0.70$) and colonization probability is $\gamma_{NGS} = 0.30$.

Assuming that enough resources are available to survey 48 sites twice each year, one might ask: is it better to have a balanced design (survey 24 sites in each habitat), or an unbalanced design in which a higher number of NGS sites are sampled (survey 20 GS sites and 28 NGS sites)? To address this question, we generated the two standard sampling designs (balanced

TABLE 6. Deviance values and likelihood ratio tests for the null hypothesis (H_0 : no habitat effect) vs. the alternative hypothesis (H_a) of habitat-specific parameter estimates under two sampling designs.

Parameter used in hypothesis	Models tested (H_0 vs. H_a)	Deviance		Test statistic		
		H_0	H_a	χ^2	df	Power
Design 1: Standard (balanced) design						
Occupancy probability	$\psi(\cdot)\epsilon(g)\gamma(g)p(\cdot,j)$ vs. $\psi(g)\epsilon(g)\gamma(g)p(\cdot,j)$	392.47	391.12	1.353	1	21%
Extinction probability	$\psi(g)\epsilon(\cdot)\gamma(g)p(\cdot,j)$ vs. $\psi(g)\epsilon(g)\gamma(g)p(\cdot,j)$	393.95	391.12	2.631	1	37%
Colonization probability	$\psi(g)\epsilon(g)\gamma(\cdot)p(\cdot,j)$ vs. $\psi(g)\epsilon(g)\gamma(g)p(\cdot,j)$	391.31	391.12	0.184	1	7%
All parameters	$\psi(\cdot)\epsilon(\cdot)\gamma(\cdot)p(\cdot,j)$ vs. $\psi(g)\epsilon(g)\gamma(g)p(\cdot,j)$	395.10	391.12	3.975	3	36%
Design 2: Unbalanced design						
Occupancy probability	$\psi(\cdot)\epsilon(g)\gamma(g)p(\cdot,j)$ vs. $\psi(g)\epsilon(g)\gamma(g)p(\cdot,j)$	389.19	387.87	1.319	1	21%
Extinction probability	$\psi(g)\epsilon(\cdot)\gamma(g)p(\cdot,j)$ vs. $\psi(g)\epsilon(g)\gamma(g)p(\cdot,j)$	388.04	387.87	2.589	1	36%
Colonization probability	$\psi(g)\epsilon(g)\gamma(\cdot)p(\cdot,j)$ vs. $\psi(g)\epsilon(g)\gamma(g)p(\cdot,j)$	390.45	387.87	0.170	1	7%
All parameters	$\psi(\cdot)\epsilon(\cdot)\gamma(\cdot)p(\cdot,j)$ vs. $\psi(g)\epsilon(g)\gamma(g)p(\cdot,j)$	391.78	387.87	3.900	3	35%

Notes: The standard balanced design has 24 sites in each of the two habitat types; the unbalanced design has 28 sites in the “poorer” amphibian habitat and 20 remaining sites in the better habitat. The alternative hypothesis (H_a) represents the true generating model $\psi(g)\epsilon(g)\gamma(g)p(\cdot,j)$, where g denotes the two habitat groups. Other terms are defined in Table 5. True differences between parameters from the different habitats are specified in the text, and all tests are based on a significance level $\alpha=0.05$. Power is approximated using methods detailed in Burnham et al. (1987).

and unbalanced) with the parameter values and detection probabilities used in previous examples: $p_1 = 0.25$ and $p_2 = 0.75$. We analyzed each design with numeric-analytic large-sample approximations (with expected values) only, under the true model used to generate the data: $\psi(g)\epsilon(g)\gamma(g)p(\cdot,j)$, where g denotes the two habitat groups. We used likelihood ratio tests to explore whether sampling designs affected the ability to determine habitat differences in initial occupancy, extinction, and colonization probabilities (Table 6). In each case the “true” generating model was considered the alternative hypothesis, and the null hypotheses were represented by reduced models containing no habitat effect for each parameter separately and for all parameters simultaneously (see Table 6 and Appendix B for a list of all candidate models). We approximated test power (assuming $\alpha = 0.05$) by using the resulting chi-square statistic as the noncentrality parameter, λ , and calculating power from a noncentral chi-squared distribution (Burnham et al. 1987).

Our results suggest very low power and little difference between the designs in the ability to detect habitat differences among model parameters (Table 6). The balanced design performed slightly better than the unbalanced design that included more sites in the “poorer” (non-geothermally influenced) habitat. Notice that the magnitudes of the differences between the habitats are not equivalent for all parameters: $|\psi_{GS} - \psi_{NGS}| = \Delta\psi = 0.20$, $|\epsilon_{GS} - \epsilon_{NGS}| = \Delta\epsilon = 0.20$, $|\gamma_{GS} - \gamma_{NGS}| = \Delta\gamma = 0.07$. Proportionally, the difference is greatest for extinction probabilities $\Delta\epsilon/\epsilon_{NGS} = 0.66$ compared to occupancy $\Delta\psi/\psi_{NGS} = 0.40$ and colonization probabilities $\Delta\gamma/\gamma_{NGS} = 0.23$. Power approximations were higher for tests involving parameters with larger proportional differences between habitats, and there was very low power to detect colonization differences in this scenario. Building on these results, investigators should be motivated to include more sites. If the number of sites were increased to 50 sites in each

habitat type, then the power to detect differences in occupancy, extinction, or all parameters simultaneously would nearly double (approximate power = 35%, 65%, and 67%, respectively). If the duration of the study were doubled for these same 100 sites (i.e., sites were sampled for eight years), then the power to detect habitat differences for extinction probability and all parameters simultaneously would rise to levels above 90%.

It is also possible to investigate the issue of power via simulation. Data are simulated under H_a , models H_a and H_0 are fit to each data set, and a likelihood ratio statistic is computed. The proportion of simulations for which H_0 is rejected is an estimate of power. Similarly, within a model selection framework, AIC_c can be computed for each simulated data set and the proportion of simulations for which H_a has the smallest AIC_c can be computed. Alternatively, the average AIC_c weights for the two models can be computed over all simulations, as another metric reflecting the discriminating ability of data resulting from a particular study design and sample size. Output from program GENPRES allows researchers to investigate all of these discriminating metrics.

DISCUSSION

Numerous investigators have emphasized the importance of clear and relevant goals when designing large-scale or long-term studies (Yoccoz et al. 2001, Pollock et al. 2002, MacKenzie et al. 2006). Well-defined study objectives are easily translated into mathematical models representing competing hypothesis about the status and behavior of the study system. Inherently, clear objectives address “why” the study is to be conducted and help determine “what” state variable is appropriate to measure. Only after the questions of “why” and “what” to sample have been adequately addressed does exploring the question of “how” to sample (i.e., survey design trade-offs) have relevance. In some cases, study design may focus on the estimation of

a set of parameters, and the objective will be to maximize precision. These parameters may include occupancy and rates of extinction and colonization, or slope parameters describing functional relationships between these basic parameters and covariates. We would hope that study design frequently focuses on discriminating among competing models of system dynamics, in which case the design objective will involve quantities such as power and discriminating ability of model selection statistics.

Program GENPRES can be used to address these kinds of design issues, for example, allowing users to explore trade-offs in temporal and spatial allocation of sample effort. Its flexibility allows scientists to tailor sampling designs to address various hypotheses and objectives, while incorporating biological and logistical constraints. Our example involving tiger salamanders in Yellowstone National Park illustrates how pilot information can be analyzed and used to inform future study designs. The available pilot data are representative of the information available on many species throughout the world, in the sense that many species have been studied within small areas compared to their overall distributions, yet these data provide a useful starting point for exploring common study design questions. If no pilot information were available, information from other systems or species, or even expert opinion could be used to explore study design trade-offs over a range of plausible parameter values. Based on the pilot data, we explored sampling designs that would allow adequate modeling of detection probability, p . We considered two multi-season sampling regimes reported in the literature (standard vs. rotating panel, McDonald 2003, MacKenzie et al. 2006), in order to determine the impact on the precision of our estimates. Finally, we explored trade-offs among designs focused on testing factors thought to influence both amphibian occupancy and vital rates.

Results from the tiger salamander example suggest that simply sampling the maximum number of sites possible, within a set of economic and logistical constraints, may not be the most advantageous design. Both single- and multiple-season scenarios revealed that occupancy and especially time-specific vital rate estimators were generally less biased under designs that include temporal survey replication both within and among seasons. The magnitude of the bias was strongly affected by the model structure for p (detection probability). In all cases, failing to model p with sufficient complexity led to severe negative bias in occupancy (single-season scenario) and vital rate estimates (multi-season studies). In the likely case in which the appropriate detection probability structure is unknown, designs involving more temporal replication at a higher proportion of sites were more robust to model misspecification. This finding is most evident when simulation-based analysis is performed, and it emphasizes the importance of investigating bias and precision with simulations for studies with small (realistic) sample sizes.

Finally, there was little difference between standard balanced vs. unbalanced study designs to detect habitat differences among multi-season model parameters. These results, together with the inference of very low power, can be extremely valuable to investigators during the study planning process. Using our balanced design, we found that doubling the sample size would certainly increase the power of detecting habitat differences in vital rates, but doubling both the sample size and study duration was necessary to increase power to >90%. This introduces yet another trade-off between the number of sites sampled per season and the number of seasons for which the study can be conducted (MacKenzie 2005, MacKenzie et al. 2006). If seasonal funding limitations prevent sampling at a large number of sites, then researchers may be required to conduct longer duration studies in order to differentiate among competing hypothesis. We also note that conclusions based upon longer duration studies are likely to be more robust to the short-term effects on the population caused, for example, by cyclic climatic conditions.

The greatest utility of program GENPRES is its flexibility to examine a wide variety of design options, tailored to a given biological system, and subject to economic constraints; thus, giving investigators an extremely useful tool during the critical planning phase of a study.

ACKNOWLEDGMENTS

We thank K. Kinkead, P. Lukacs, and two anonymous reviewers for comments that substantially improved the paper. We also thank D. Patla, C. Peterson, P. S. Corn, R. Klaver, and R. Bennetts for data and discussion about amphibians in the Greater Yellowstone National Park Service Network. This work as supported, in part, by the USGS Northeast Amphibian Research and Monitoring Initiative.

LITERATURE CITED

- Barbraud, C., J. D. Nichols, J. E. Hines, and H. Hafner. 2003. Estimating rates of local extinction and colonization in colonial species and an extension to the metapopulation and community levels. *Oikos* 101:113–126.
- Burnham, K. P., D. R. Anderson, G. C. White, C. Brownie, and K. P. Pollock. 1987. Design and analysis of methods for fish survival experiments based on release–recapture. *American Fisheries Society Monograph* 5:1–437.
- Cochran, W. G. 1977. Sampling techniques. John Wiley, New York, New York, USA.
- Corn, P. S., B. R. Hossack, E. Muths, D. A. Patla, C. R. Peterson, and A. L. Gallant. 2005. Status of amphibians on the Continental Divide: surveys on a transect from Montana to Colorado, USA. *Alytes* 22:85–94.
- Field, S. A., A. J. Tyre, and H. P. Possingham. 2005. Optimizing allocation of monitoring effort under economic and observational constraints. *Journal of Wildlife Management* 69:473–482.
- Fisher, R. N., and H. B. Shaffer. 1996. The decline of amphibians in California's Great Central Valley. *Conservation Biology* 10:1387–1397.
- Gimenez, O., A. Viallefont, E. A. Catchpole, R. Choquet, and B. J. T. Morgan. 2004. Methods for investigating parameter redundancy. *Animal Biodiversity and Conservation* 27:561–572.

- Gu, W., and R. K. Swihart. 2004. Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. *Biological Conservation* 116:195–203.
- Hames, R. S., K. V. Rosenberg, J. D. Lowe, and A. A. Dhondt. 2001. Site reoccupancy in fragmented landscapes: testing predictions of metapopulation theory. *Journal of Animal Ecology* 70:182–190.
- Koch, E. D., and C. R. Peterson. 1995. Amphibians and reptiles of Yellowstone and Grand Tetons National Parks. University of Utah Press, Salt Lake City, Utah, USA.
- MacKenzie, D. I. 2005. What are the issues with “presence/absence” data for wildlife managers? *Journal of Wildlife Management* 69:849–860.
- MacKenzie, D. I., J. D. Nichols, J. E. Hines, M. G. Knutson, and A. D. Franklin. 2003. Estimating site occupancy, colonization and local extinction when a species is detected imperfectly. *Ecology* 84:2200–2207.
- 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.
- MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. P. Pollock, L. L. Bailey, and J. E. Hines. 2006. Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence. Academic Press, San Diego, California, USA.
- MacKenzie, D. I., J. D. Nichols, N. Sutton, K. Kawanishi, and L. L. Bailey. 2005. Improving inference in population studies of rare species that are detected imperfectly. *Ecology* 86:1101–1113.
- MacKenzie, D. I., and J. A. Royle. 2005. Designing efficient occupancy studies: general advice and tips on allocation of survey effort. *Journal of Applied Ecology* 42:1105–1114.
- Manley, P. N., W. J. Zielinski, M. D. Schlesinger, and S. R. Mori. 2004. Evaluation of a multiple-species approach to monitoring species at the ecoregional scale. *Ecological Applications* 14:296–310.
- Martinez-Solano, I., J. Bosch, and M. Garcia-Paris. 2003. Demographic trends and community stability in a montane amphibian assemblage. *Conservation Biology* 17:238–244.
- Mazerolle, M. J., A. Desrochers, and L. Rochefort. 2005. Landscape characteristics influence pond occupancy by frogs after accounting for detectability. *Ecological Applications* 15:824–834.
- McDonald, T. L. 2003. Review of environmental monitoring methods: survey designs. *Environmental Monitoring and Assessment* 85:277–292.
- Moilanen, A. 2002. Implications of empirical data quality to metapopulation model parameter estimation and application. *Oikos* 96:516–530.
- Muths, E., R. E. Jung, L. L. Bailey, M. J. Adams, P. S. Corn, C. K. Dodd, Jr., G. M. Fellers, W. J. Sadinski, C. R. Schwalbe, S. C. Walls, R. N. Fisher, A. L. Gallant, and W. A. Battaglin. 2005. The U.S. Department of Interior’s Amphibian Research and Monitoring Initiative (ARM): a successful start to a national program. *Applied Herpetology* 2:355–371.
- Nichols, J. D., B. R. Noon, S. L. Stokes, and J. E. Hines. 1981. Remarks on the use of mark-recapture methodology in estimating avian population size. Pages 121–136 in C. J. Ralph and M. J. Scott, editors. *Estimating the numbers of terrestrial birds*. Studies in Avian Biology. Volume 6. Cooper Ornithological Society, Camarillo, California, USA.
- O’Connell, A. F., N. W. Talancy, L. L. Bailey, J. R. Sauer, R. Cook, and A. T. Gilbert. 2006. Estimating site occupancy and detection probability parameters for meso- and large mammals in a coastal ecosystem. *Journal of Wildlife Management* 70, *in press*.
- Pollock, K. H., J. D. Nichols, T. R. Simons, and J. R. Sauer. 2002. Large scale wildlife monitoring studies: statistical methods for design and analysis. *Environmetrics* 13:1–15.
- Scott, J. M., P. J. Heglund, M. L. Morrison, J. B. Hauffer, M. G. Raphael, W. A. Wall, and F. B. Samson, editors. 2002. *Predicting species occurrences: issues of accuracy and scale*. Island Press, Washington, D.C., USA.
- Tyre, A. J., B. Tenhumberg, S. A. Field, D. Niejalke, K. Parris, and H. P. Possingham. 2003. Improving precision and reducing bias in biological surveys: estimating false-negative error rates. *Ecological Applications* 13:1790–1801.
- Van Buskirk, J. 2005. Local and landscape influence on amphibian occurrence and abundance. *Ecology* 86:1936–1947.
- White, G. C., and K. P. Burnham. 1999. Program MARK: survival estimation from populations of marked animals. *Bird Study* 46:120–139.
- Wintle, B. A., M. A. McCarthy, K. M. Parris, and M. A. Burgman. 2004. Precision and bias of methods for estimating point survey detection probabilities. *Ecological Applications* 14:703–712.
- Yoccoz, N. G., J. D. Nichols, and T. Boulinier. 2001. Monitoring of biological diversity in space and time. *Trends in Ecology and Evolution* 16:446–453.

APPENDIX A

User-defined code for single-season models that are not in predefined list (*Ecological Archives* A017-010-A1).

APPENDIX B

List of candidate models for exploration of habitat differences (*Ecological Archives* A017-010-A2).