EFFICIENT ESTIMATION OF ABUNDANCE FOR PATCHILY DISTRIBUTED POPULATIONS VIA TWO-PHASE, ADAPTIVE SAMPLING

MICHAEL J. CONROY,1,5 JONATHAN P. RUNGE,2,6 RICHARD J. BARKER,3 MATTHEW R. SCHOFIELD,3,7 AND CHRISTOPHER J. FONNESBECK4,8

1U.S. Geological Survey, Georgia Cooperative Fish and Wildlife Research Unit, Warnell School of Forestry and Natural Resources, University of Georgia, Athens, Georgia 30602 USA
2Warnell School of Forestry and Natural Resources, University of Georgia, Athens, Georgia 30602 USA
3Department of Mathematics and Statistics, University of Otago, P.O. Box 56, Dunedin, New Zealand
4Florida Fish and Wildlife Research Institute, St. Petersburg, Florida 33701 USA

Abstract. Many organisms are patchily distributed, with some patches occupied at high density, others at lower densities, and others not occupied. Estimation of overall abundance can be difficult and is inefficient via intensive approaches such as capture–mark–recapture (CMR) or distance sampling. We propose a two-phase sampling scheme and model in a Bayesian framework to estimate abundance for patchily distributed populations. In the first phase, occupancy is estimated by binomial detection samples taken on all selected sites, where selection may be of all sites available, or a random sample of sites. Detection can be by visual surveys, detection of sign, physical captures, or other approach. At the second phase, if a detection threshold is achieved, CMR or other intensive sampling is conducted via standard procedures (grids or webs) to estimate abundance. Detection and CMR data are then used in a joint likelihood to model probability of detection in the occupancy sample via an abundance–detection model. CMR modeling is used to estimate abundance for the abundance–detection relationship, which in turn is used to predict abundance at the remaining sites, where only detection data are collected. We present a full Bayesian modeling treatment of this problem, in which posterior inference on abundance and other parameters (detection, capture probability) is obtained under a variety of assumptions about spatial and individual sources of heterogeneity. We apply the approach to abundance estimation for two species of voles (Microtus spp.) in Montana, USA. We also use a simulation study to evaluate the frequentist properties of our procedure given known patterns in abundance and detection among sites as well as design criteria. For most population characteristics and designs considered, bias and mean-square error (MSE) were low, and coverage of true parameter values by Bayesian credibility intervals was near nominal. Our two-phase, adaptive approach allows efficient estimation of abundance of rare and patchily distributed species and is particularly appropriate when sampling in all patches is impossible, but a global estimate of abundance is required.

Key words: abundance estimation; adaptive sampling; Bayesian estimation; capture–mark–recapture; occupancy estimation; two-phase sampling.

INTRODUCTION

Abundance is a central driver of many ecological processes, and estimating abundance is a common problem in ecological statistics. Accurate estimates of abundance lead to better understanding of competition, density dependence, and species range patterns, among other processes. Additionally, estimating abundance for rare organisms has become an important goal given the increasing effects humans have on their environment (Thompson 2004). Thus, obtaining accurate estimates of abundance remains an important task for ecologists in general, with crucial consequences for endangered species.

Problems associated with estimating abundance of rare species over large areas are well-documented (Thompson 2004). The essential problem involves patchily distributed organisms, with some patches occupied at high density, others at lower densities, and still others not at all. Estimation of abundance for the total population in all patches can be difficult. It is certainly inefficient to exert equal amounts of sampling effort (e.g., via capture–mark–recapture [CMR]) in all patches. Other approaches for sampling rare species have been proposed including adaptive sampling (Thompson 1992, Manly et al. 2002a, Manly 2004) for clustered populations, consideration of patch occupancy rather than abundance as the state variable of interest in...
order to incorporate a larger spatial extent (MacKenzie et al. 2002, 2006), unequal probability sampling combined with resource selection functions (Manly et al. 2002b, McDonald 2004), and line-transect designs (Becker et al. 2004). The approach proposed here can incorporate all of these while providing rigorous estimates of both patch occupancy and abundance.

The sampling design we propose has two steps. In the first step, occupancy sampling (MacKenzie et al. 2006) occurs and in the second, a subset of those patches found to be occupied is sampled for abundance. Thus, we propose a form of two-phase sampling (Cochran 1977, Thompson 1992) to estimate both occupancy and abundance. A similar method has recently been proposed (Royle et al. 2007), but ours differs because its two-phase nature allows for unequal probability of selection among sample patches and thus the ability to sample larger spatial extent at lower cost.

This method will have broad application for ecological sampling but is especially well suited to organisms that are low in abundance, patchily distributed, and that can be sampled intensively (i.e., marked and recaptured or resighted on multiple occasions), such as Canada lynx (Lynx canadensis), Alabama beach mouse (Peromyscus poliotonus annobates), northern leopard frog (Lithobates pipiens), gold striped gecko (Hoplodactylus crysosireticus), and Mead’s milkweed (Asclepias meadiii).

**Sampling overview**

Simple occupancy models assume that detection probabilities are homogeneous among sites. Heterogeneity in detection may be accommodated by finite and continuous mixture distributions (MacKenzie et al. 2006, Royle 2006) or by modeling detection probability as a function of site-specific abundance (Royle and Nichols 2003, MacKenzie et al. 2006, Royle 2006). In the latter case, under certain (e.g., Poisson) assumptions about the spatial distribution of animals, estimates of abundance may be derived, but in general these models treat abundance as random effect, possibly modeled as a function of covariates (Royle 2006, Royle and Nichols 2003). Presumably, if abundance were independently observable at a sample of sites, it would be possible to use abundance values directly as a covariate in abundance–detection models, and thereby to predict abundance in the remaining sites.

Here we propose a sampling scheme and model in which two-phase sampling is employed in a Bayesian framework to estimate abundance for patchily distributed populations. In the first phase, occupancy is estimated via k binomial detection samples on each of m sites, resulting in yi = 0, 1, 2, ..., k detections per site. Initially we assume that m = M (the total number of sites available). This phase allows estimation of rates of occupancy and numbers of patches occupied, taking into account possibly heterogeneous probability of detection via k detection samples per patch, from which are recorded the number of samples y for which a detection of >0 animals occurs. Detection can be by visual surveys, detection of sign, physical captures, or other approach. The only requirements being (1) positive determination of a detection, and (2) the k samples are independent, Bernoulli trials (e.g., over time or among independent observers).

At the second phase, if a detection threshold (τ) is achieved (yi > τ) capture–mark–recapture (CMR) sampling is conducted via standard procedures (e.g., a 100 × 100 m trapping grid) for t sampling occasions, and typical closed-population CMR models (Otis et al. 1978, Williams et al. 2002) are used to estimate site-specific abundance, N. For example, τ = 0 indicates that only one detection would trigger CMR sampling. The detection and CMR data are then used in a joint likelihood to model probability of detection in the occupancy sample via an abundance–detection model. The CMR model is then used to estimate abundance at the m1 sites where CMR sampling is conducted and to estimate the abundance–detection relationship. Finally, the abundance–detection model is used to predict abundance at the remaining m1 – m sites.

**Sampling design**

For the first phase of sampling we take k1 detection samples on each of i = 1, ... m study sites, where initially we assume that the m = M constitute the set of all possible habitat patches that potentially are occupied by the population. We later generalize this to allow for probability sampling of m < M. Detections may be via any method; methods for small mammals include trapping, searches for tracks, scat, or other sign, or detection via tracking tubes. In each of the j = 1, ... k1 detection samples at site i we record xij = 1 if one or more animals are detected, and xij = 0 otherwise, and summarize the outcomes as

\[ y_i = \sum_{j=1}^{k_1} x_{ij}; \quad y_i = 0, 1, 2, \ldots, k_1. \]

At the second phase, on each of the m1 sites for which a detection threshold yi > τ is achieved, we conduct intensive sampling (e.g., CMR sampling) in order to estimate abundance. The CMR samples are taken over l = 1, ..., ti, capture occasions, and we record the resulting capture histories X. For convenience, but without loss of generality, we assume that detection and CMR sampling have equal effort across sites (i.e., ki = k; ti = l). Initially we assume that CMR samples are taken at all sites for which any detections occur (τ = 0). Later we generalize the procedure to allow for higher detection thresholds (τ ≤ k) and for probability sampling of sites achieving a specified threshold.

**Statistical model**

The statistical model is based on the joint distribution of the detection data \{yi\}, i = 1, ..., n and the capture histories X on the j = 1, ..., m1 sites for which, initially, yi
> 0 (τ = 0). We model the processes leading to the detection and the capture histories as conditionally independent:
\[
[y_i|d_i][X_i|N_i, p_i] = \prod_{i=1}^{m} \left( \binom{k}{y_i} d_i^{y_i} (1 - d_i)^{k-y_i} \right) \times \prod_{j=1}^{m} [X_j|N_j, p_j]
\]
where \(d_i\) is the probability of detection on the occupancy samples, \(p_i\) is a vector of probability of capture in the CMR samples at site \(i\), and the notation \([Y|X]\) is used to denote the unspecified probability density function for the random variable \(Y\) conditional on \(X\). The first component of this product is equivalent to a zero-inflated binomial model allowing for heterogeneous probability of detection (MacKenzie et al. 2006). Following Royle and Nichols (2003) we parameterized \(d_i\) as a function of per-site abundance:
\[
d_i = 1 - (1 - r)^{N_i}
\]
where \(r\) is the probability that an individual is detected in the occupancy samples if present, and is assumed constant among animals and over sites. In addition to modeling heterogeneity in detection probabilities \(d_i\), Eq. 2 provides a basis for prediction of abundance on sites on the \(m - m_1\) sites on which intensive sampling does not occur, as detailed below. The second component specifies a general relationship between observed capture histories and per-site abundance \(N_i\) and probability of capture \(p_i\), with \(p_i\) potentially varying over capture occasions, among animals, or as a function of previous capture (Otis et al. 1978, Williams et al. 2002). Initially we specify a simple model for the CMR process, in which \(p\) is allowed to vary among sites, but is constant otherwise \((p_i = p)\); this is equivalent to specification of model \(M_0\) (Otis et al. 1978) stratified by sites. We later generalize the modeling of \(p\) via a log-linear approach, allowing for time, behavioral, and other forms of heterogeneity (see Appendix).

The full Bayesian model additionally requires a model specifying variation in site-specific abundance \(\{N_i\}\), as well as prior distributions for all parameters. Many possibilities exist for modeling \(\{N_i\}\), potentially involving spatial autocorrelation and the incorporation of covariates. For simplicity, we modeled abundance using a gamma mixture of Poisson distributions:
\[
[N_i] \sim \gamma(\lambda_i), \quad \{\lambda_i\} \sim \{\lambda_i\} | [\alpha, \beta] \sim \gamma(\alpha, \beta)
\]
where
\[
[N_i] \sim P(\{\lambda_i\}), \quad [\lambda_i] | [\alpha, \beta] \sim \gamma(\alpha, \beta).
\]
The relationship in Eq. 3 results in \(N_i\), the marginal distribution of abundance, following a negative binomial distribution, with heterogeneity in local density \(\lambda_i\) controlled by the parameters \(\alpha\) and \(\beta\). Specification of prior parameter distributions results in a joint model for the parameters and data:
\[
\prod_{i=1}^{m} [y_i|r, N_i, k][\lambda_i|\alpha, \beta][X_i|N_i, p_i][N_i|\lambda_i][p_i]|r[\alpha|\beta]
\]
\[
= \prod_{i=1}^{m} \left( \frac{k}{y_i} \right) \left( \frac{1 - (1 - r)^{N_i}}{\left(1 - (1 - r)^{1 - r}N_i\right)^y} \right) \left( \frac{\lambda_i}{\alpha, \beta} \right) \prod_{j=1}^{m} [X_j|N_j, p_j]|r[\alpha|\beta]
\]
where \([r|p, \alpha|\beta]\) is the joint prior distribution for per-animal detection, capture probabilities, and the Poisson-gamma parameters. Primary interest is on site-specific abundance, and aggregations of abundance across sites, where
\[
N_i^{tot} = \sum_{i=1}^{m_i} N_i
\]
is total abundance at the sites selected for CMR sampling,
\[
N_2^{tot} = \sum_{i=m_1+1}^{m} N_i
\]
is abundance predicted at the \(m - m_1\) non-selected sites, and
\[
N_i^{tot} = N_1^{tot} + N_2^{tot}
\]
is abundance on all \(m\) sites.

Although we were primarily interested in estimation of abundance, we note that the model in Eq. 4 can incorporate site occupancy as a state variable, by re-expressing the first term of the product as
\[
\prod_{i=1}^{m} \left( \frac{k}{y_i} \right) \left( \frac{1 - (1 - r)^{N_i}}{\left(1 - (1 - r)^{1 - r}N_i\right)^y} \right) \left( 1 - Z_i \right) I(N_i = 0) \left( \frac{\lambda_i}{\alpha, \beta} \right)
\]
where \(Z_i = I(N_i > 0)\) specifies whether site \(i\) is occupied (1 or not 0). Our model for \(\{N_i\}\) can then be used to obtain the probability of occupancy as
\[
\psi_i = \Pr(Z_i = 1) = 1 - \Pr(N_i = 0).
\]
The model in Eqs. 3–5 effectively allows simultaneous estimation of abundance on the CMR sites, estimation of occupancy, and prediction of abundance on sites for which CMR samples are not taken but detections are observed, via the relationships in Eqs. 2 and 3.

We envisage at least two special cases of the model:
1. Retention of joint detection-CMR data, homogeneous detection \((d_i = d)\).—In this case, the detections \(y_i\) provide no information on abundance on non-CMR sites, other than obviously \(y_i > 0\) implies \(N_i > 0\). Modeling of abundance on the CMR sites then depends heavily on the assumed abundance model (Eq. 3). Essentially, abundance on the non-CMR sites is predicted from the observed detections.
2. Detection only data, heterogeneous detection under $d_i = 1 - (1 - r)^N_i$.—Under this case, no direct information on abundance is available from any sites, and inference on abundance follows from Eq. 2 and restrictive (e.g., Poisson) assumptions about the distribution of abundance among site. This leads to the Royle-Nichols model (Royle and Nichols 2003).

In both special cases, and our more general case, it is assumed that the same process generates abundance on all sites in the study. This has implications for study design, as discussed later.

Case study

We applied our approach to data for two species of vole (*Microtus pennsylvanicus* and *M. montanus*) trapped in Montana, USA (Runge 2005, Runge et al. 2007). As part of this study, four trap grids were randomly established and sampled over five occasions to estimate grid-specific abundance. Each grid consisted of 0.8 ha trapped in each of two habitats. For the purposes of this study, we treat each habitat within each grid as a “site” ($m = 8$ sites). We treated these data as a complete sample, from which subsamples were taken according to our adaptive procedure. First, we randomly selected a single row of 10 traps from each site, and used the 10 traps from the first night of trapping as our occupancy sample ($y_i$). On each site for which $y_i > \tau$, we used the observed five-night capture histories on all rows to estimate abundance, and estimated occupancy and abundance for the entire study area. Because of the relatively small number of sites and potential detections, we consider only two thresholds: $\tau = 0, 1$. Analysis of these data using programs MARK (White and Burnham 1999) and CAPTURE (White et al. 1982) for all sites suggested similar capture probabilities among sites, but evidence of behavioral or individual heterogeneity within site. We therefore estimated parameters for models incorporating time, behavioral, individual, spatial, or no heterogeneity in capture probabilities, for a total of eight CMR models (see Appendix).

We implemented this model using program WinBUGS (Lunn et al. 2000), a freely available program specialized for fitting hierarchical, Bayesian models such as ours. We used U(0, 1) priors for parameters such as $r$, $p$, and $d$ that were on the probability scale, N(0, 0.001) for coefficients of the log-linear CMR models, and Ga(0.001, 0.001) for $\alpha$ and $\beta$, the parameters controlling site-to-site variation in density. For each analysis, we ran two independent chains for 100,000 iterations, discarding the first half for burn-in, examining plots of model traces and Gelman-Rubin statistics to confirm convergence (Gelman et al. 2004). Annotated code and data are provided in Supplement 1.

Simulation study

We used a simulation study to evaluate the frequentist properties of our procedure under a range of known patterns in abundance and detection among sites, and design criteria. First, we specified an $m$-dimensioned vector $\mathbf{N}$, where $m$ is the total number of sites in the study area, as a negative binomial variate:

$$N_i = \text{NB}(r, p)$$

where $r$ and $p$ were obtained via moments of the distribution as $p = \bar{x}/\bar{X^2}$, $r = \bar{x}p/(1 - p)$. We selected $\bar{x} = 100$ and $\bar{X^2} = 100, 1000$, and 10,000, providing a range of densities among sites to nearly “random” (i.e., Poisson-distributed) to very over-dispersed; we would expect the latter in cases where abundance is very patchily distributed among sites. Conditioning on the vector of $\mathbf{N}$, we then generated the observed states of detection and capture as follows.

For each site, we simulated the number of detections $y_i$ in $k$ detection samples as a binomial variate:

$$y_i \sim B(k, d_i) I(N_i > 0)$$

where

$$d_i = 1 - (1 - r)^N_i$$

thus specifying that per-sample detection probability is a function of site abundance, $N_i$. We then simulated capture histories for a $t$-occasion CMR experiment based on the $N_i$ individuals occurring in each site where $y_i > \tau$. Because our interest centers on the tradeoff between detection and CMR sampling components, and not on CMR modeling per se, we used a simple model to generate capture histories, assuming probability of capture ($p$) is the same for all individuals and at all times. We considered small to moderately large studies ($m = 5, 10, 25$, and 50, $k = 5$ and 10, $t = 5$ and 10); per-individual detection of $r = 0.0005$, 0.005; capture probabilities of $p = 0.5$; and detection thresholds $\tau = -1$ (i.e., no threshold, all sites had CMR sampling), 0, 1, and 3. Finally, in addition to the above factors, we considered one set of trials where the sampling fraction at the second (CMR) phase varied from $\xi = 0.25$ (one-quarter of sites where $y_i > \tau$) to 1 (all sites where $y_i > \tau$ samples, as above). Because of the enormous number of combinations of these factors, we fixed most factors equal to values in the above ranges. Exceptions included the factors under investigation (above), and $\bar{X^2}$, which we set at 100000 because we were most interested in design behavior in patchily distributed populations. We also considered two cases of extremely overdispersed populations having $\bar{X^2} = 50000$ and 100000, and explored combinations of design factors $\tau$, $\xi$, and $k$ for these populations. We avoided certain factor ranges as these led to uninteresting situations, e.g., $r > 0.01$ results in nearly perfect detection of occupancy, $\bar{X^2} < 1000$ leads to all sites being occupied with few extreme abundance values.

For each simulation we obtained 20,000 Markov Chain Monte Carlo (MCMC) samples, discarding the first 10,000 for burn-in; based on preliminary runs this number of MCMC samples was sufficient for convergence, given the simplicity of the model (i.e., no CMR heterogeneity). Data were simulated using Python,
We summarized statistics for combined detection–CMR (capture–mark–recapture) modeling of abundance for *Microtus* spp. in Montana, USA.

<table>
<thead>
<tr>
<th>Sites</th>
<th>(y_i)</th>
<th>(n_i)</th>
<th>(u_i)</th>
<th>(M_i)</th>
<th>(m_i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>237</td>
<td>119</td>
<td>359</td>
<td>118</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>114</td>
<td>46</td>
<td>127</td>
<td>68</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>65</td>
<td>29</td>
<td>80</td>
<td>36</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>52†</td>
<td>22†</td>
<td>58†</td>
<td>30†</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>100</td>
<td>47</td>
<td>133</td>
<td>53</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>96</td>
<td>36</td>
<td>109</td>
<td>60</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Notes: Definitions of variables: \(y_i\) is the number of traps (out of 10 total) that captured Microtus for the occupancy sample; \(n_i\) is the total number of trapping occurrences in the five-night capture session; \(u_i\) is the total number of new captures over the five-night capture session; \(M_i\) is the number of marked animals available for capture at each occasion, summed over all five occasions; and \(m_i\) is the total number of recaptures over the five-night capture session.

† Excluded when \(\tau = 1\), where \(\tau\) is the detection threshold (e.g., \(\tau = 0\) indicates that only one detection would trigger CMR sampling).

†† CMR statistics do not exist for these sites because the detection threshold of \(y > 0\) was not achieved.

formatted for WinBUGS (program available online). Implementation of the MCMC process was as above for the *Microtus* example, except that we had but a single, homogeneous-capture CMR model, with \(p\) assumed constant among sites, for which we specified a U(0, 1) prior.

We repeated the simulation–MCMC process 500 times for each combination of \(\hat{s}^2, m, k, t, r, p, \tau\), and \(\xi\) and evaluated bias, precision, and interval coverage for all model parameters. We evaluated relative bias (RBIAS) and mean-square error (MSE) as

\[
\text{RBIAS} = \frac{1}{\ell} \sum_{i=1}^{m} (\hat{\theta}_i - \theta_0)
\]

and

\[
\text{RMSE} = \sqrt{\frac{1}{\ell} \sum_{i=1}^{m} (\hat{\theta}_i - \theta_0)^2}
\]

where \(\theta_0\) is the value of the parameter of interest (\(N_{1\text{tot}}, N_{2\text{tot}}, \text{or} N_{\text{tot}}\)) value at the \(i\)th simulation trial and \(\hat{\theta}_0\) is the posterior mean from the MCMC samples for that parameter. Finally, we evaluated the performance of the Bayesian credible intervals (BCI) as confidence intervals by computing the proportion of 95% BCIs in \(\ell\) simulation trials that included \(\theta_0\) in the interval.

Annotated Python and WinBugs code for the simulations is provided in Supplement 2.

RESULTS

Data example

Mice were detected on six of eight sites, with detection frequencies ranging from one to six of the \(k = 10\) detection trials per site (Table 1). On the sites where the detection threshold was achieved, we obtained estimates of \(p, r,\) and \(N_{1\text{tot}}\) under models with either homogeneous or site-specific capture probabilities and permitting time, individual, behavioral, or no heterogeneity in capture probabilities. Goodness of fit and model selection criteria strongly favored a model allowing individual heterogeneity in capture probabilities, specific to each site, for either threshold criterion (Table 2). We used this model to predict abundance on the remaining sites for each of the two detection thresholds (\(\tau = 0, 1\); Table 3). Parameter estimates were similar, with lower precision when fewer CMR sites were included (\(\tau = 1\)).

We note that Runge (2005) conducted more extensive trapping on the last two grids and failed to capture any animals, suggesting that overall abundance was close to the "naïve" estimates of \(N_{1\text{tot}}\), similar to the results for these areas from program MARK. Thus, for this example, \(N_{2\text{tot}}\) may have been close to 0 for \(\tau = 0\), and thus \(N_{\text{tot}}\) overestimated. For the case when \(\tau = 1\), there was a minimum of 22 animals present (\(u_2 = 22\)) on the site not included in the CMR sample, and this value is easily included in the 95% BCI for \(N_{1\text{tot}}\) under either model. Of course, in practice we would not know the true abundance states of any of the sites.

Simulation study

We computed bias, MSE, and coverage for all parameters, but present here only the parameters of primary interest, \(N_{1\text{tot}}, N_{2\text{tot}},\) and \(N_{\text{tot}}\). For most combinations of population characteristics and design criteria, estimates of model parameters had low bias and MSE and interval coverage was close to the nominal 95% (Table 4). Some loss of accuracy occurred, as expected, with increasing detection threshold \(\tau\) and decreasing sample proportion \(\xi\), with performance relatively poor for \(\xi = 0.25\). However, under extreme variability (\(s^2 = 50000, 100000\)) in abundance, bias and MSE were high...
and coverage low, depending on design. For these conditions, under designs in which no threshold was applied, but a random sample of CMR sites selected, bias and MSE were comparatively low and coverage was near nominal. We discuss implications of these results for future study designs in Discussion.

**DISCUSSION**

Our example and simulation study suggest that an adaptive design and integrated Bayesian analytical approach can provide efficient abundance estimation for patchily distributed animal populations. The approach appears to work well over a broad range of conditions and design factors. The approach appears particularly well suited to situations where abundance is a state variable of key interest and where conventional approaches to sampling abundance, such as CMR grids, are highly inefficient due to high variation in local densities. By invoking a Bayesian hierarchical approach, we are able to use an adaptive design to provide inference on abundance on all sites, and thus estimates of abundance throughout the study area.

Although we have illustrated our detection-abundance sampling scheme using CMR for the abundance component, the scheme is readily extended to other data structures, such as distance sampling or trapping webs (Buckland et al. 2001, Williams et al. 2002). Because the detection and abundance components of our data model

<table>
<thead>
<tr>
<th>CMR model†</th>
<th>Deviance (−2 log-likelihood)</th>
<th>Goodness of fit</th>
<th>Bayes factor‡</th>
<th>No. parameters</th>
<th>ΔAIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>τ = 0§</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(M_0(s))</td>
<td>603.7</td>
<td>0.051</td>
<td>(4.22645 \times 10^{-16})</td>
<td>5</td>
<td>63.5</td>
</tr>
<tr>
<td>(M_1(s))</td>
<td>580.9</td>
<td>0.433</td>
<td>(3.77513 \times 10^{-11})</td>
<td>15</td>
<td>60.7</td>
</tr>
<tr>
<td>(M_2(s))</td>
<td>613.2</td>
<td>0.358</td>
<td>(3.65659 \times 10^{-18})</td>
<td>8</td>
<td>70.0</td>
</tr>
<tr>
<td>(M_3(s))</td>
<td>602.8</td>
<td>0.365</td>
<td>(6.62839 \times 10^{-16})</td>
<td>33</td>
<td>118.6</td>
</tr>
<tr>
<td>(M_4(s))</td>
<td>540.2</td>
<td>0.251</td>
<td>(0.025991129)</td>
<td>5</td>
<td>0.0</td>
</tr>
<tr>
<td>(M_5(s))</td>
<td>532.9</td>
<td>0.647</td>
<td>1</td>
<td>15</td>
<td>12.7</td>
</tr>
<tr>
<td>(M_6(s))</td>
<td>612.3</td>
<td>0.135</td>
<td>(5.73468 \times 10^{-18})</td>
<td>4</td>
<td>70.1</td>
</tr>
<tr>
<td>(M_7(s))</td>
<td>598.3</td>
<td>0.3104</td>
<td>(6.28884 \times 10^{-15})</td>
<td>9</td>
<td>66.1</td>
</tr>
<tr>
<td>τ = 1§</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(M_8(s))</td>
<td>540.1</td>
<td>0.05162</td>
<td>(5.71987 \times 10^{-17})</td>
<td>5</td>
<td>68.4</td>
</tr>
<tr>
<td>(M_9(s))</td>
<td>518.5</td>
<td>0.3724</td>
<td>(2.80393 \times 10^{-12})</td>
<td>13</td>
<td>62.8</td>
</tr>
<tr>
<td>(M_{10}(s))</td>
<td>546.1</td>
<td>0.2124</td>
<td>(2.84776 \times 10^{-18})</td>
<td>8</td>
<td>80.4</td>
</tr>
<tr>
<td>(M_{11}(s))</td>
<td>538.5</td>
<td>0.7235</td>
<td>(1.27298 \times 10^{-16})</td>
<td>28</td>
<td>112.8</td>
</tr>
<tr>
<td>(M_{12}(s))</td>
<td>471.7</td>
<td>0.2952</td>
<td>(0.040762204)</td>
<td>5</td>
<td>0.0</td>
</tr>
<tr>
<td>(M_{13}(s))</td>
<td>465.3</td>
<td>0.6075</td>
<td>1</td>
<td>13</td>
<td>9.6</td>
</tr>
<tr>
<td>(M_{14}(s))</td>
<td>547</td>
<td>0.1265</td>
<td>(1.81581 \times 10^{-18})</td>
<td>4</td>
<td>73.3</td>
</tr>
<tr>
<td>(M_{15}(s))</td>
<td>532.9</td>
<td>0.2853</td>
<td>(2.09337 \times 10^{-15})</td>
<td>8</td>
<td>67.2</td>
</tr>
</tbody>
</table>

† In models, 0 denotes homogeneity within site, b denotes behavioral response, h denotes individual heterogeneity, t denotes time variation, s denotes variation among sites, and a “dot” (·) denotes absence of variation among sites in capture probability.

‡ Ratio of likelihood under given model to that under the model with the highest likelihood (lowest deviance).

§ Detection threshold (e.g., \(\tau = 0\) indicates that only one detection would trigger CMR sampling).

Table 2. Model selection statistics for combined detection–CMR modeling of abundance for *Microtus* spp. in Montana, USA.

<table>
<thead>
<tr>
<th>CMR model</th>
<th>Parameter</th>
<th>Median</th>
<th>95% Bayesian credible interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Lower</td>
<td>Upper</td>
</tr>
<tr>
<td>(\tau = 0)†</td>
<td>(N_{1}^{tot})</td>
<td>305.1</td>
<td>300.7</td>
</tr>
<tr>
<td>(\tau = 0)†</td>
<td>(N_{2}^{tot})</td>
<td>27.68</td>
<td>2.79</td>
</tr>
<tr>
<td>(\tau = 0)†</td>
<td>(N_{3}^{tot})</td>
<td>334.6</td>
<td>307.1</td>
</tr>
<tr>
<td>(\tau = 1)‡</td>
<td>(N_{1}^{tot})</td>
<td>280.1</td>
<td>277.9</td>
</tr>
<tr>
<td>(\tau = 1)‡</td>
<td>(N_{2}^{tot})</td>
<td>50.32</td>
<td>10.49</td>
</tr>
<tr>
<td>(\tau = 1)‡</td>
<td>(N_{3}^{tot})</td>
<td>331.0</td>
<td>290.8</td>
</tr>
</tbody>
</table>

Note: \(N_{m}^{tot}\) is total abundance at the sites selected for CMR sampling, \(N_{m}^{tot}\) is abundance predicted at the \(m - m_0\), non-selected sites, and \(N_{s}^{tot}\) is abundance on all \(s\) sites.

‡ Detection threshold (e.g., \(\tau = 0\) indicates that only one detection would trigger CMR sampling).
Realized state and sampled outcomes

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Factors</th>
<th>$Z_{tot}$ (Range)</th>
<th>$N_i$ (Max.)</th>
<th>$m_i$ (Range)</th>
<th>RBIAS</th>
<th>RMSE</th>
<th>Coverage</th>
<th>RBIAS</th>
<th>RMSE</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario 1</td>
<td>$\tau = -1$</td>
<td>22–25</td>
<td>376</td>
<td>25–25</td>
<td>0.002</td>
<td>0.007</td>
<td>0.934</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\tau = 0$</td>
<td>23–25</td>
<td>379</td>
<td>13–25</td>
<td>0.001</td>
<td>0.007</td>
<td>0.944</td>
<td>0.047</td>
<td>0.548</td>
<td>0.960</td>
</tr>
<tr>
<td></td>
<td>$\tau = 1$</td>
<td>22–25</td>
<td>387</td>
<td>10–24</td>
<td>0.000</td>
<td>0.007</td>
<td>0.924</td>
<td>0.071</td>
<td>0.367</td>
<td>0.956</td>
</tr>
<tr>
<td></td>
<td>$\tau = 2$</td>
<td>23–25</td>
<td>376</td>
<td>3–18</td>
<td>0.001</td>
<td>0.008</td>
<td>0.948</td>
<td>0.051</td>
<td>0.263</td>
<td>0.950</td>
</tr>
<tr>
<td>Scenario 2</td>
<td>$\tau = 0$, $\xi = 0.25$</td>
<td>22–25</td>
<td>383</td>
<td>1–13</td>
<td>0.001</td>
<td>0.015</td>
<td>0.934</td>
<td>0.050</td>
<td>0.390</td>
<td>0.888</td>
</tr>
<tr>
<td></td>
<td>$\tau = 0$, $\xi = 0.5$</td>
<td>23–25</td>
<td>375</td>
<td>5–19</td>
<td>0.000</td>
<td>0.010</td>
<td>0.934</td>
<td>0.017</td>
<td>0.217</td>
<td>0.946</td>
</tr>
<tr>
<td></td>
<td>$\tau = 0$, $\xi = 0.75$</td>
<td>22–25</td>
<td>391</td>
<td>7–23</td>
<td>0.000</td>
<td>0.008</td>
<td>0.952</td>
<td>0.022</td>
<td>0.239</td>
<td>0.954</td>
</tr>
<tr>
<td></td>
<td>$\tau = 0$, $\xi = 1$</td>
<td>23–25</td>
<td>379</td>
<td>13–25</td>
<td>0.001</td>
<td>0.007</td>
<td>0.944</td>
<td>0.047</td>
<td>0.548</td>
<td>0.960</td>
</tr>
<tr>
<td></td>
<td>$\tau = -1$, $\xi = 0.5$</td>
<td>23–25</td>
<td>373</td>
<td>6–20</td>
<td>0.002</td>
<td>0.010</td>
<td>0.934</td>
<td>0.027</td>
<td>0.219</td>
<td>0.948</td>
</tr>
<tr>
<td>Scenario 3</td>
<td>$k = 5$, $r = 5$</td>
<td>22–25</td>
<td>381</td>
<td>12–23</td>
<td>0.001</td>
<td>0.007</td>
<td>0.940</td>
<td>0.097</td>
<td>0.528</td>
<td>0.932</td>
</tr>
<tr>
<td></td>
<td>$k = 10$, $r = 5$</td>
<td>23–25</td>
<td>379</td>
<td>13–25</td>
<td>0.001</td>
<td>0.007</td>
<td>0.944</td>
<td>0.047</td>
<td>0.548</td>
<td>0.960</td>
</tr>
<tr>
<td></td>
<td>$k = 5$, $r = 10$</td>
<td>23–25</td>
<td>383</td>
<td>11–24</td>
<td>0.001</td>
<td>0.015</td>
<td>0.582</td>
<td>0.105</td>
<td>0.526</td>
<td>0.942</td>
</tr>
<tr>
<td></td>
<td>$k = 10$, $r = 20$</td>
<td>22–25</td>
<td>383</td>
<td>14–25</td>
<td>0.002</td>
<td>0.003</td>
<td>0.525</td>
<td>0.055</td>
<td>0.594</td>
<td>0.944</td>
</tr>
<tr>
<td>Scenario 4</td>
<td>$s^2 = 10000$</td>
<td>23–25</td>
<td>379</td>
<td>13–25</td>
<td>0.001</td>
<td>0.007</td>
<td>0.944</td>
<td>0.047</td>
<td>0.548</td>
<td>0.960</td>
</tr>
<tr>
<td></td>
<td>$s^2 = 100000$</td>
<td>5–21</td>
<td>1126</td>
<td>1–17</td>
<td>0.000</td>
<td>0.006</td>
<td>0.960</td>
<td>4.219</td>
<td>4.782</td>
<td>0.172</td>
</tr>
<tr>
<td></td>
<td>$s^2 = 50000$</td>
<td>10–24</td>
<td>809</td>
<td>5–18</td>
<td>0.001</td>
<td>0.018</td>
<td>0.816</td>
<td>1.852</td>
<td>2.178</td>
<td>0.486</td>
</tr>
<tr>
<td>Scenario 5</td>
<td>$r = 0.0005$</td>
<td>22–25</td>
<td>380</td>
<td>2–15</td>
<td>0.001</td>
<td>0.009</td>
<td>0.932</td>
<td>0.195</td>
<td>0.555</td>
<td>0.900</td>
</tr>
<tr>
<td></td>
<td>$r = 0.005$</td>
<td>23–25</td>
<td>379</td>
<td>13–25</td>
<td>0.001</td>
<td>0.007</td>
<td>0.944</td>
<td>0.047</td>
<td>0.548</td>
<td>0.960</td>
</tr>
<tr>
<td>Scenario 6</td>
<td>$m = 10$</td>
<td>9–10</td>
<td>288</td>
<td>4–10</td>
<td>0.002</td>
<td>0.011</td>
<td>0.951</td>
<td>0.082</td>
<td>0.930</td>
<td>0.941</td>
</tr>
<tr>
<td></td>
<td>$m = 25$</td>
<td>23–25</td>
<td>379</td>
<td>13–25</td>
<td>0.001</td>
<td>0.007</td>
<td>0.944</td>
<td>0.047</td>
<td>0.548</td>
<td>0.960</td>
</tr>
<tr>
<td></td>
<td>$m = 50$</td>
<td>47–50</td>
<td>444</td>
<td>33–48</td>
<td>0.001</td>
<td>0.005</td>
<td>0.950</td>
<td>0.055</td>
<td>0.390</td>
<td>0.942</td>
</tr>
<tr>
<td></td>
<td>$m = 100$</td>
<td>95–100</td>
<td>518</td>
<td>72–93</td>
<td>0.001</td>
<td>0.004</td>
<td>0.948</td>
<td>0.072</td>
<td>0.297</td>
<td>0.928</td>
</tr>
<tr>
<td>Scenario 7</td>
<td>$K = 20$, $\tau = 0$, $\xi = 1$</td>
<td>10–24</td>
<td>833</td>
<td>5–21</td>
<td>0.001</td>
<td>0.007</td>
<td>0.942</td>
<td>1.758</td>
<td>1.941</td>
<td>0.438</td>
</tr>
<tr>
<td></td>
<td>$K = 10$, $\tau = -1$, $\xi = 0.5$</td>
<td>10–23</td>
<td>884</td>
<td>4–20</td>
<td>0.003</td>
<td>0.010</td>
<td>0.924</td>
<td>0.095</td>
<td>0.703</td>
<td>0.916</td>
</tr>
<tr>
<td></td>
<td>$K = 20$, $\tau = -1$, $\xi = 0.5$</td>
<td>11–25</td>
<td>828</td>
<td>5–19</td>
<td>0.002</td>
<td>0.010</td>
<td>0.926</td>
<td>0.082</td>
<td>0.420</td>
<td>0.942</td>
</tr>
</tbody>
</table>

Notes: Definitions of variables: $Z_{tot}$, total number of occupied sites; $N_i$, abundance on site $i$, $i = 1, \ldots, n$; RBIAS, relative bias; RMSE, relative mean-square error; $m_i$, no. CMR sites sampled; $m$, no. sites in study population; $s^2$, population variance of per-site abundance; $r$, per-animal detection probability; $\tau$, detection threshold; $\xi$, proportion of sites meeting threshold that are sampled; $k$, no. detection samples per site; $t$, no. CMR occasions per sample.

We agree that under certain assumptions (e.g., proportionality between local and superpopulation abundance) (Eq. 1) are conditionally independent, the second (abundance) component could simply be replaced by the appropriate abundance model, with the estimation of the detection–abundance relationship and prediction of abundance on non-sampled sites. Similarly, another potentially useful generalization is to relax the assumption of closure between the detection sample and the intensive sample. This approach might be needed in large-scale surveys where it is not possible to conduct all sampling over short time intervals where closure can be assumed. Of course, such a generalization would require the estimation of additional parameters (i.e., survival, recruitment, movement) to allow dynamic modeling of abundance, perhaps via a two-phase extension of the robust design for CMR estimation (Pollock 1982, Kendall and Pollock 1992). An additional point, raised by reviewer J. D. Nichols, is that certain conditions (e.g., intensive sampling over one day with occupancy sampling over several days) might allow separate estimation of abundance and detection conditional on local availability, and abundance and detection in the “superpopulation” (sensu Kendall 1999) potentially available for capture over the duration of the study. We agree that under certain assumptions (e.g., proportionality between local and superpopulation abundance)
this might improve joint estimation, and believe that further work in this area could be fruitful.

Our simulation studies indicate relative insensitivity to a wide range of population conditions and design factors, but some patterns emerge that have bearing on study design. In particular, an extremely over-dispersed population potentially results in a situation where most detections would occur on the fewer, high-density sites. In this case, using a threshold detection criterion for CMR or other abundance sampling may lead to poor prediction of abundance on the non-CMR sites. We suggest that in such situations, or when in doubt, it may be preferable not to apply a threshold detection criterion, but rather to randomly select the CMR samples from among all sites (from which detection samples are also taken), in this way ensuring a better representation of inter-site densities in both samples.

Finally, as suggested by the above discussion, proper application of this approach requires at least provisional knowledge of site-specific abundance distributions, detection rates, capture rates, and other factors. Application of this approach would benefit from the acquisition of data on these factors from pilot studies, which can then be used, together with predictive models such as the ones we have describe, to develop optimal sampling schemes.

We envisage that this approach to estimating abundance will have broad applicability across several areas of ecological research, and complements existing approaches (e.g., MacKenzie et al. 2006, Royle et al. 2007). Our two-phase, adaptive approach should allow estimation of rare and patchily distributed species in a more cost-effective manner than conventional designs. Our approach can also aid in investigating more basic approaches to ecology such as abundance–range patterns because it enables statistically rigorous estimation of abundance over large areas. Thus the model proposed will be widely adaptable to many ecological investigations.

ACKNOWLEDGMENTS

This work was initiated while the first author was a William Evans visiting fellow at the University of Otago. We thank D. I. MacKenzie and M. Efford for useful ideas in the early phases of this work, and J. D. Nichols, J. T. Peterson, J. A. Royle, and an anonymous referee for helpful comments on earlier drafts. The Georgia Cooperative Fish and Wildlife Research Unit is jointly sponsored by USGS, U.S. Fish and Wildlife Service, the University of Georgia, Georgia Department of Natural Resources, and the Wildlife Management Institute.

LITERATURE CITED


and modeling. Elsevier-Academic, San Diego, California, USA.


APPENDIX

Modeling of capture-mark-recapture (CMR) data for joint detection-CMR modeling of occupancy and abundance (Ecological Archives E089-192-A1).

SUPPLEMENT 1

WinBUGS model code for analysis of Microtus example (Ecological Archives E089-192-S1).

SUPPLEMENT 2

Python and WinBUGS model code for simulation study (Ecological Archives E089-192-S2).