Methods in Ecology and Evolution

Methods in Ecology and Evolution 2016, **7**, 285–293 doi: 10.1111/2041-210X.12499

Occupancy estimation for rare species using a spatially-adaptive sampling design

Krishna Pacifici[1]*, Brian J. Reich[2], Robert M. Dorazio[3] and Michael J. Conroy[4]

[1]*Department of Applied Ecology, North Carolina State University, Raleigh, NC 27695, USA;* [2]*Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA;* [3]*Southeast Ecological Science Center, U.S. Geological Survey Gainesville, FL 32653, USA; and* [4]*Warnell School of Forestry and Natural Resources, University of Georgia, Athens, GA 30602, USA*

Summary

1. Spatially clustered populations create unique challenges for conservation monitoring programmes. Advances in methodology typically are focused on either the design or the modelling stage of the study but do not involve integration of both.

2. We integrate adaptive cluster sampling and spatial occupancy modelling by developing two models to handle the dependence induced by cluster sampling. We compare these models to scenarios using simple random sampling and traditional occupancy models via simulation and data collected on a rare plant species, *Tamarix ramosissima*, found in China.

3. Our simulations show a marked improvement in confidence interval coverage for the new models combined with cluster sampling compared to simple random sampling and traditional occupancy models, with greatest improvement in the presence of low detection probability and spatial correlation in occupancy.

4. Accounting for the design using the simple cluster random-effects model reduces bias considerably, and full spatial modelling reduces bias further, especially for large n when the spatial covariance parameters can be estimated reliably. Both new models build on the strength of occupancy modelling and adaptive sampling and perform at least as well, and often better, than occupancy modelling alone.

5. We believe our approach is unique and potentially useful for a variety of studies directed at patchily distributed, clustered or rare species exhibiting spatial variation.

Key-words: adaptive cluster sampling, informative sampling, probit regression, rare species, spatial regression, *Tamarix ramosissima*

Introduction

Monitoring state variables such as animal abundance, density or site occupancy rate is a critical component of many large-scale and long-term conservation efforts. Unfortunately, this information can be challenging to obtain due to limited financial resources and logistical constraints imposed by complex environmental and geographic conditions (Possingham *et al.* 2001). These problems are exacerbated when the species of interest is patchily distributed, occurs in low numbers or is difficult to observe or capture (Thompson 2004; MacKenzie *et al.* 2005; Martin *et al.* 2014). These characteristics, possessed by many rare or elusive species, create unique challenges for both the design and analysis of surveys (Thompson 2004; Cunningham & Lindenmayer 2005; MacKenzie *et al.* 2005; Pacifici, Dorazio & Conroy 2012).

Recent advances in methodology have focused on either improving sampling design or developing sophisticated data analysis. Several approaches have been proposed that tailor data collection to profit from a specific behaviour or characteristic of the species of interest. For instance, the use of designs such as stratified sampling (Edwards *et al.* 2005), sequential sampling (Thompson 2002, 2004), multiphase sampling (Thompson 2002, 2004; Pacifici, Dorazio & Conroy 2012) or adaptive sampling (Thompson 1990, 2004; Brown *et al.* 2013) can potentially increase the information content in a particular sample as well as provide more efficient estimation by accounting for spatial structure in species distributions.

Other approaches have been developed to increase the flexibility of estimation models. Recent advancements add model complexity to account for deficiencies in the design or data. For example, site occupancy models (MacKenzie *et al.* 2002; Tyre *et al.* 2003; Bailey, MacKenzie & Nichols 2014) can now account for lack of independence among observations collected on trails or transects (Hines *et al.* 2010; Aing *et al.* 2011; Guillera-Arroita *et al.* 2011, 2012) and spatial autocorrelation (Hoeting, Leecaster & Bowden 2000; Hooten, Larsen & Wikle 2003; Royle & Dorazio 2008; Gardner *et al.* 2010; Johnson *et al.* 2013). Other approaches have been developed to permit inference at multiple spatial scales (Nichols *et al.* 2008), or to leverage information across different species (MacKenzie *et al.* 2005; Alldredge *et al.* 2007), or community characteristics (Dorazio *et al.* 2006; Zipkin, DeWan & Royle 2009; Dorazio, Gotelli & Ellison 2011; Pacifici *et al.* 2014).

*Correspondence author. E-mail: jkpacifi@ncsu.edu

Although a focus on either the design or modelling stage can be useful, ideally both should be considered to maximize the quality of information collected and rigour of inference (e.g. Johnson, Laake & Ver Hoef 2010). This was recognized early on by Thompson (Adaptive cluster sampling; 1990) and MacKenzie *et al.* (occupancy modelling; 2006), but has not seen much attention in the literature. The dual focus on applying advanced designs and modelling techniques should, theoretically, be most beneficial in cases where one technique is not sufficient due to a major impediment. For example, the reason to resort to more sophisticated designs is often to ensure sufficient data collection even for simple analyses, especially for species that are rare, hard to detect, or exhibit strong spatial correlation (Pacifici, Dorazio & Conroy 2012). Conversely, a major impediment to fitting more complicated models is a lack of data forcing researchers to focus on simpler study objectives and models.

MacKenzie *et al.* (2006) and Thompson (1990) both recognized that allocating more effort in areas conditional on a previous detection (augmenting the sample design) can potentially increase the performance of the estimators. They describe but did not develop a scenario that would require the combination of occupancy estimation (MacKenzie *et al.* 2006) and adaptive cluster sampling (Thompson 1990) and point out the potential advantages of such an approach. MacKenzie & Royle (2005) also suggested the possibility of selecting sites by adaptive sampling leading to reliable inference about occupancy probability although they did not develop the model in detail. Conroy *et al.* (2008) presented an approach that relied on detections to augment the sample with additional information in the form of capture–recapture. Rapley & Welsh (2008) developed a model-based approach for adaptive cluster sampling that includes the benefits of both approaches, but does not incorporate estimates of detection probability or measurement error therefore making reliable inference difficult. Peyrard *et al.* (2013) also addressed spatial adaptive sampling, but were more concerned in reconstructing a spatial map of the system and not estimating potential covariates that influence the spatial distribution of a species.

In this paper, we develop a statistically rigorous approach that integrates adaptive cluster sampling with a spatially explicit occupancy analysis. The adaptive cluster sample begins with a simple random sample and is followed by a second stage of additional sampling in areas that are occupied in the first stage. Inclusion of the second stage should increase the precision of species distribution maps in occupied patches and provide information about local spatial variation in occupancy to improve inference via spatial analysis. We develop two models based on data collected in an adaptive manner: a computationally demanding spatial model based on a latent continuous spatial process and a simpler version with shared random effects. Our adaptive sampling scheme for rare species is designed to put more sampling effort in areas likely to be occupied, however, we show that by analysing data using our proposed spatial model, we avoid bias that may occur due to preferential sampling. We evaluate our approach using both simulated and real data and find that the proposed sampling and analysis method is more powerful than simple random sampling for rare species with low detection rates.

Materials and Methods

ADAPTIVE SAMPLING SCHEME

The sampling begins with J sampling occasions at each of n spatial locations selected randomly from the spatial domain of interest. Denote s_{i1} and $Y_{i1} \in \{0, \ldots, J\}$ as the spatial location and number of detections, respectively, for sampling site $i = 1,\ldots, n$. Additional sampling effort is dedicated to areas near locations s_{i1} with detections $Y_{i1} > 0$. For each location with a detection, an additional $k_i - 1$ sites $s_{i2}, \ldots, s_{ik_i}$ near s_{i1} are surveyed, again with J sampling occasions at each site. Note that each selected sample location is surveyed the same number of times, even if the same location is selected more than once by adaptive cluster sampling. We allocate each site to the first cluster it is assigned to and resolve any issues with individual sites appearing in multiple clusters. This process could be repeated further, but for simplicity we consider only a two-stage sampling design.

There are several potential ways to draw the second-stage site locations. We partition the spatial domain into a fine rectangular grid of cells, sample n cells in the first stage, and then the four rook neighbours (neighbours to the north, east, south and west) of occupied sites in the second stage (with each cell attributed to at most one cluster). Another possibility is to randomly sample second-stage sites within a certain radius of the sites that were occupied in the first stage. It is also not required that J sampling occasions be dedicated to each sampling site, but we assume this for simplicity.

STATISTICAL MODEL

We use a state-space approach in which we express the model by its two component processes: a sub-model for the latent occupancy state and a sub-model for the observations conditional on the latent occupancy state. The true occupancy state is Z_{ij}, where $Z_{ij} = 1$ indicates the site is occupied and $Z_{ij} = 0$ indicates the site is not occupied for $i = 1,\ldots, n$ initial sites and $j = 2, \ldots, 5$ second-stage sites (here we only look at the four rooks neighbouring sites). Conditioned on the occupancy state the observations are modelled as

$$Y_{ij}|Z_{ij}, P \sim \text{Binomial}(J, pZ_{ij}) \qquad \text{eqn 1}$$

where P is the conditional probability of detection at an occupied site. Thus, if a site is occupied, then the data are binomial with J trials and success probability P, otherwise $Prob\left(Y_{ij} = 0\right) = 1$. We assume the detection probability is constant, although it can be modelled using site-level or replicate-level covariates (MacKenzie *et al.* 2002). We consider both a spatial and non-spatial model for the occupancy states Z_{ij}, as described below.

SPATIAL OCCUPANCY MODEL

We assume that there is a latent continuous spatial process V_{ij} so that the site is occupied ($Z_{ij} = 1$) if $V_{ij} > 0$, and the site is not occupied ($Z = 0$) if $V_{ij} \leq 0$. The latent continuous process is modelled as a Gaussian process with mean dependent on covariates X_{ij}, $E\left(V_{ij}\right) = X_{ij}^T\beta$, variance $\text{Var}(V_{ij}) = 1$ and spatial correlation $\text{Cor}(V_{ij}, V_{uv}) = M\left(\|s_{ij} - s_{uv}\|\right)$, where $(\|s_{ij} - s_{uv}\|$ is the distance between the sites and M is the Matern correlation function (Cressie 1993). The Matern correlation function has two parameters: $\eta > 0$ controls the smoothness of V, and $\rho > 0$ controls the range of spatial dependence. Under this model, the occupancy probability at site s_{ij} has the usual probit regression form $\Phi(X_{ij}^T\beta)$, where Φ is the standard normal distribution function (see Dorazio & Rodríguez 2012; for an application of occupancy modelling with probit regression). We use a probit link because updating the latent parameters can be made with standard Gibbs sampling and this avoids having to select tuning parameters for these updating steps (Johnson *et al.* 2013; Data S1). Predictions at unsampled locations are made using standard Bayesian kriging methods as described in the supplement.

RANDOM-EFFECTS OCCUPANCY MODEL

In addition to the spatial model for the latent occupancy indicators, we use a simpler model which accounts for dependence between sites in the same sampling cluster using a shared random effect. In this model, the latent process is modelled as

$$V_{ij} = X_{ij}^T\beta + b_i + \varepsilon_{ij} \qquad \text{eqn 2}$$

where the cluster random effects are $b_i \sim$ Normal(0, σ^2) and $\varepsilon_{ij} \sim$ Normal(0, 1), both independent over i and j. To illustrate how the shared random effect accounts for dependence within a cluster, we observe that the correlation between V_{ij} and V_{uv} is zero if they are from different clusters ($i \neq u$) and $\sigma^2/(\sigma^2 + 1)$ if the observations are from the same cluster ($i = u$). Therefore, when σ^2 is small, all observations are independent, and when σ^2 is large, observations within a cluster are highly correlated. The motivation for this dependence structure is that the random effects can account for local dependence within a cluster. Unlike the spatial model, nearby clusters are assumed to be independent, which may be reasonable if the clusters are sufficiently separated. For prediction at unsampled locations, we treat each cell as its own cluster and draw b_i and V_{ij} from (2). We note that other clustering schemes are possible for the unsampled locations, but clustering does not affect the marginal occupancy probability.

INFORMATIVE SAMPLING

Our sampling scheme is geared towards increasing sampling effort in areas likely to be occupied. The subsequent statistical analysis then requires care to ensure estimates are not biased by informative sampling (Diggle, Menezes & Su 2010; Pati, Reich & Dunson 2011). Below, we argue that because site selection is completely determined by observed data and not unknown parameters, the statistical model does not need to account for informative sampling. We use bracket notation (Gelfand & Smith 1990) to specify probability density functions; thus, $[x, y]$ denotes the joint density of random variables X and Y, $[x|y]$ denotes the conditional density of X given $Y = y$, and $[x]$ denotes the unconditional (marginal) density of X.

Let θ be the collection of unknown parameters including the true occupancy status of each spatial location, detection probability, regression coefficients and spatial correlation parameters. The sampling has two stages. Denote s_1 as the initial set of n locations sampled, and s_2 as the second set of locations sampled in clusters around sites with positive results. At these two sets of sampling locations, data y_1 and y_2 are collected. The preferential sampling literature (Diggle, Menezes & Su 2010) handles informative sampling by treating the location of samples s_1 and s_2 as random variables. The entire posterior is then

$$[\theta|s_1, s_2, y_1, y_2] \propto [s_1, s_2, y_1, y_2|\theta][\theta],$$

where $[s_1, s_2, y_1, y_2|\theta]$ is the likelihood of the data given the parameters and $[\theta]$ is the prior. Following the sequential nature of the sampling, the likelihood becomes

$$[s_1, s_2, y_1, y_2|\theta] = [s_1|\theta][y_1|s_1, \theta][s_2|s_1, y_1, \theta][y_2|s_1, s_2, y_1, \theta]$$

The selection of the initial sites s_1 is completely random and does not depend on the parameters, and the selection of the second set of points s_2 depends only on s_1 and y_1 and not the parameters, therefore $[s_1|\theta] = [s_1]$ and $[s_2|s_1, y_1, \theta] = [s_2|s_1, y_1]$, and thus

$$[\theta|s_1, s_2, y_1, y_2] \propto [y_1|s_1, \theta][y_2|s_1, s_2, y_1, \theta][\theta] = [y_1, y_2|s_1, s_2, \theta][\theta]$$

This is the usual spatial regression model for the data y_1 and y_2 that ignores site selection. Therefore, because site selection depends only on observed data and not unknown parameters, we can proceed with the standard statistical methods.

BAYESIAN ANALYSIS

To complete the Bayesian model, we specify priors for the hyperparameters. For all models, we assume the detection probability has prior $p \sim$ Unif(0, 1) and the elements of $\boldsymbol{\beta}$ have independent normal priors with mean 0 and variance 100. For the spatial model, the spatial correlation parameters have log-normal priors $\eta \sim$ LN(0, 1) and $\rho \sim$ LN(0, 10). For the cluster random-effects model, the variance has inverse gamma prior $\sigma^2 \sim$ Inv$G$(0.1, 0.1). We develop specific MCMC samplers (Supplement 1) to update all parameters in both models using R (R Development Core Team 2014). We generate 10 000 samples and discard the first 2 000 as burn-in. Convergence is monitored using traces plots of several representative parameters.

Simulation study

DESIGN

In this section, we compare our new method's ability to estimate the proportion of the spatial domain that is occupied. We compare two different sampling designs: adaptive cluster sampling (Clus) and simple random sampling (SRS) along with three different models of analysis: standard occupancy model (In – for independent) (MacKenzie *et al.* 2002), shared random-effects model (RE) and the spatial occupancy model (Sp). We generate data on a 20-by-20 rectangular grid of $N = 400$ locations. The data are generated from the spatial occupancy model described above. The true occupancy proportion is $\psi = \sum_l^N Z^l/N$ where Z^l is the occupancy state for cell l and ψ is the mean over the N cells in the spatial domain. The true occupancy proportion is controlled by the parameter β, where $E(V_l) = \Phi(\beta)$. We set β so that the expected value of ψ is either 0·1 or 0·4. The spatial correlation parameters are set to $\eta = 2$ and $\rho = 0.5$ and the detection probability was set to either $p = 0{\cdot}25$ or $0{\cdot}75$. We ran simulations with $\psi = 0.01$, but there were too few data in the majority of cases with enough detections (~99% out of 500 in some cases) so we did not report these results because of this selective bias and subsequent biased results.

The spatially-adaptive design (Clu) samples n sites in the first stage and in the second stage the four rook neighbours of the first-stage sites with at least one occurrence. We simulate data with n = 20, 50, 100 and 150 initial sites. We compare this with data collected using a simple random sample (SRS) with the same total number of sites that were sampled in the adaptive design ($\sum_{i=1}^n k_i$). All methods used $J = 5$ sampling occasions for all sites. For each data set gathered following the adaptive design, we fit the spatial occupancy model (Sp) and cluster random-effects model (RE) described above, as well as the independence model (In) with independent V_{ij} (i.e. the spatial model with $\eta = 0$). For the data gathered following a simple random sample, we fit both the spatial (Sp) and independent models (In).

For each combination of β, p and n, we simulated 500 data sets; data sets with less than 3 occurrences for either sampling design were discarded (this occurred for less than 10% of data sets for all scenarios we considered). For each sampling design and analytic method, we computed the posterior mean and posterior 90% interval of the occupancy proportion ψ. Tables 1-3 give the bias, mean squared error (MSE) and coverage for ψ.

Results

As expected, gathering data using the clustered design yet analysing the data without regard to the sampling design using the independence model results in positive bias for ψ (Table 1). This occurs because more sampling effort is dedicated to locations that are likely occupied, leading to overestimation using naive methods. Accounting for the design using the simple cluster random-effects model reduces bias considerably, and full spatial modelling reduces bias further, especially for large n when the spatial covariance parameters can be estimated reliably. Similar problems are observed with the independence model with coverage far below the nominal level (Table 2). The independence model's poor coverage is likely the result of both bias and underestimating uncertainty because potential duplication (spatial dependence) between nearby observations is ignored. For the spatial model, coverage of 90% intervals is at least 86·8% for all settings with $n > 20$.

Mean squared error values indicated that cluster sampling is advantageous for rare and elusive species when sample sizes, occurrence and detection rates are low (Table 3). For the case with true occupancy $\psi = 0.1$ and detection $p = 0.25$, the cluster sample/spatial model approach has smaller MSE (up to 50% reduction) than the simple random sampling designs. In this case, the cluster random-effects model also provides a substantial reduction in MSE compared to the simple random sampling methods and is therefore a viable alternative to the full spatial model. In fact, with $n = 20$, the simpler random-effects model actually has smaller MSE than the full spatial model. In other cases with either high occupancy ($\psi = 0.4$) or detection ($p = 0.75$), the simple random sampling method with spatial analysis give smaller MSE than the adaptive sampling methods. However, even in these cases, the cluster sampling method with spatial analysis has comparable bias and coverage near the nominal level. Final sample sizes after adaptive sampling were higher when true occupancy and detection were both high and in some cases represented ~71% increase in sampling effort (Table S1).

Data example

In addition to simulations, we illustrate our new methods on a population of plant species, *Tamarix ramosissima*, from a 1-km^2 study area in the Inner Mongolia region of PR China (Smith *et al.* 2012). The Chinese Academy of Forestry counted all desert shrubs within the study area, and therefore, we obtained a complete census including the number of stems and distribution at the time of the study from which we could subsample following the different survey designs (Clu and SRS). Density of *Tamarix* was 0·15 m^{-2} and occupancy was 6% of 10-m × 10-m grid cells. In addition, we incorporated several derived habitat variables including elevation, slope, distance to roads, solar insolation and topographic convergence. These variables were derived using the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Digital Elevation Model (DEM) from NASA Jet Propulsion labora-

Table 1. Bias (Monte Carlo error) from the simulation study. Data are sampled using either cluster ('Clu') or simple random sampling ('SRS'), and data are analysed using either spatial ('Sp'), random-effects ('RE'), or independent ('In') models with $J = 5$ sampling occasions. The simulations vary by the expected occupancy proportion (ψ), detection probability (p) and initial sample size (n)

ψ	p	n	Clu/Sp	Clu/RE	Clu/In	SRS/Sp	SRS/In
0·1	0·25	20	0·184 (0·03)	0·130 (0·02)	0·280 (0·03)	0·214 (0·03)	0·218 (0·03)
		50	0·050 (0·01)	0·045 (0·01)	0·130 (0·02)	0·084 (0·01)	0·098 (0·02)
		100	0·022 (0·01)	0·020 (0·01)	0·079 (0·01)	0·029 (0·01)	0·051 (0·01)
		150	0·011 (0·01)	0·009 (0·01)	0·052 (0·01)	0·019 (0·01)	0·025 (0·01)
	0·75	20	0·040 (0·00)	−0·016 (0·00)	0·098 (0·01)	0·014 (0·00)	0·014 (0·00)
		50	0·013 (0·00)	−0·026 (0·00)	0·069 (0·00)	0·000 (0·00)	0·001 (0·00)
		100	0·005 (0·00)	−0·020 (0·00)	0·051 (0·00)	0·000 (0·00)	0·001 (0·00)
		150	0·002 (0·00)	−0·013 (0·00)	0·039 (0·00)	0·001 (0·00)	0·002 (0·00)
0·4	0·25	20	0·158 (0·01)	0·104 (0·01)	0·260 (0·01)	0·121 (0·01)	0·134 (0·01)
		50	0·032 (0·00)	0·054 (0·01)	0·147 (0·01)	0·022 (0·00)	0·047 (0·01)
		100	0·010 (0·00)	0·049 (0·00)	0·102 (0·00)	0·004 (0·00)	0·013 (0·00)
		150	0·007 (0·00)	0·049 (0·00)	0·081 (0·00)	0·008 (0·00)	0·013 (0·00)
	0·75	20	0·020 (0·00)	−0·001 (0·01)	0·122 (0·00)	0·003 (0·00)	0·005 (0·00)
		50	0·010 (0·00)	0·022 (0·00)	0·117 (0·00)	0·006 (0·00)	0·006 (0·00)
		100	0·008 (0·00)	0·039 (0·01)	0·098 (0·00)	−0·001 (0·00)	0·001 (0·00)
		150	0·002 (0·00)	0·036 (0·00)	0·078 (0·00)	0·001 (0·00)	0·001 (0·00)

Table 2. Coverage of 90% intervals (Monte Carlo error) from the simulation study. Data are sampled using either cluster ('Clu') or simple random sampling ('SRS'), and data are analysed using either spatial ('Sp'), random-effects ('RE') or independent ('In') models with $J = 5$ sampling occasions. The simulations vary by the expected occupancy proportion (ψ), detection probability (p) and initial sample size (n)

ψ	p	n	Clu/Sp	Clu/RE	Clu/In	SRS/Sp	SRS/In
0·1	0·25	20	0·803 (0·04)	0·855 (0·04)	0·575 (0·05)	0·808 (0·04)	0·793 (0·04)
		50	0·864 (0·02)	0·867 (0·02)	0·586 (0·03)	0·883 (0·02)	0·878 (0·02)
		100	0·884 (0·02)	0·823 (0·02)	0·568 (0·02)	0·901 (0·02)	0·884 (0·02)
		150	0·880 (0·01)	0·831 (0·02)	0·557 (0·02)	0·880 (0·01)	0·880 (0·02)
	0·75	20	0·888 (0·02)	0·935 (0·01)	0·547 (0·03)	0·935 (0·01)	0·935 (0·01)
		50	0·905 (0·02)	0·789 (0·02)	0·414 (0·02)	0·898 (0·01)	0·902 (0·01)
		100	0·880 (0·01)	0·635 (0·02)	0·334 (0·02)	0·870 (0·01)	0·873 (0·01)
		150	0·876 (0·02)	0·619 (0·02)	0·298 (0·02)	0·839 (0·02)	0·829 (0·02)
0·4	0·25	20	0·798 (0·02)	0·838 (0·02)	0·525 (0·02)	0·848 (0·02)	0·826 (0·02)
		50	0·880 (0·01)	0·832 (0·02)	0·460 (0·02)	0·896 (0·01)	0·850 (0·02)
		100	0·870 (0·02)	0·780 (0·02)	0·364 (0·02)	0·904 (0·01)	0·894 (0·01)
		150	0·872 (0·02)	0·717 (0·02)	0·332 (0·03)	0·886 (0·02)	0·857 (0·02)
	0·75	20	0·868 (0·02)	0·854 (0·02)	0·424 (0·02)	0·896 (0·01)	0·914 (0·01)
		50	0·886 (0·02)	0·814 (0·02)	0·150 (0·02)	0·906 (0·01)	0·901 (0·01)
		100	0·880 (0·02)	0·733 (0·02)	0·027 (0·02)	0·907 (0·02)	0·933 (0·02)
		150	0·902 (0·02)	0·505 (0·03)	0·010 (0·01)	0·875 (0·02)	0·854 (0·02)

Table 3. Mean squared error (Monte Carlo error) from the simulation study. Data are sampled using either cluster ('Clu') or simple random sampling ("SRS"), and data are analysed using either spatial ('Sp'), random-effects ('RE') or independent ('In') models with $J = 5$ sampling occasions. The simulations vary by the expected occupancy proportion (ψ), detection probability (p) and initial sample size (n)

ψ	p	n	Clu/Sp	Clu/RE	Clu/In	SRS/Sp	SRS/In
0·1	0·25	20	0·081 (0·02)	0·048 (0·01)	0·139 (0·02)	0·101 (0·02)	0·105 (0·02)
		50	0·015 (0·01)	0·018 (0·01)	0·040 (0·01)	0·039 (0·01)	0·046 (0·01)
		100	0·006 (0·01)	0·009 (0·01)	0·019 (0·01)	0·010 (0·01)	0·023 (0·01)
		150	0·003 (0·00)	0·003 (0·00)	0·008 (0·01)	0·008 (0·00)	0·009 (0·01)
	0·75	20	0·005 (0·00)	0·003 (0·00)	0·016 (0·00)	0·003 (0·00)	0·003 (0·00)
		50	0·001 (0·00)	0·002 (0·00)	0·007 (0·00)	0·001 (0·00)	0·001 (0·00)
		100	0·000 (0·00)	0·001 (0·00)	0·004 (0·00)	0·000 (0·00)	0·001 (0·00)
		150	0·000 (0·00)	0·000 (0·00)	0·002 (0·00)	0·000 (0·00)	0·000 (0·00)
0·4	0·25	20	0·068 (0·00)	0·042 (0·00)	0·109 (0·01)	0·052 (0·00)	0·057 (0·00)
		50	0·010 (0·00)	0·016 (0·00)	0·036 (0·00)	0·007 (0·00)	0·017 (0·00)
		100	0·003 (0·00)	0·008 (0·00)	0·014 (0·00)	0·002 (0·00)	0·003 (0·00)
		150	0·002 (0·00)	0·006 (0·00)	0·009 (0·00)	0·001 (0·00)	0·002 (0·00)
	0·75	20	0·008 (0·00)	0·016 (0·00)	0·023 (0·00)	0·004 (0·00)	0·004 (0·00)
		50	0·002 (0·00)	0·007 (0·00)	0·016 (0·00)	0·001 (0·00)	0·002 (0·00)
		100	0·001 (0·00)	0·004 (0·00)	0·011 (0·00)	0·000 (0·00)	0·000 (0·00)
		150	0·000 (0·00)	0·002 (0·00)	0·006 (0·00)	0·000 (0·00)	0·000 (0·00)

tory and resampled to a 10 m × 10 m cell resolution to match the scale of the species and habitat variables (see Appendix 17·1 in Smith *et al.* 2012 for more details).

Taking the occupancy status in the observed data as the true value of Z^l (Fig. 1, top left panel), we simulated data following the same procedure as in the simulation study. We subsample the original data using either $n = 100$, 250 or 500 initial locations for the cluster sampling approach (Clu), and compare with an equally sized simple random sample (SRS). This provides the true occurrence state at those locations wherein we simulate detection/non-detection data to use in our model comparisons. The detection probability was taken to be either $p = 0{\cdot}25$ or 0·75. For each value of n and p we generated 500 data sets. The models used to analyse each simulated data set were the same as those in the simulation study (Sp, RE, In), except that the covariates in the mean (X_{ij}) now include DEM-derived habitat variables (elevation, slope, distance to roads, solar insolation and topographic convergence). The effect of these habitat variables was generated from the estimated effect from the original data (Smith *et al.* 2012). We used the same priors for the parameters as in the simulation study. For each fitted model, we computed bias and MSE of the posterior mean occupancy proportion ψ. In addition, to compare accuracy of species distribution maps, we compute the posterior occupancy probability in each cell, $\hat{Z}^l$, and the Brier score (Gneiting & Raftery 2007) $\sum_{l=1}^{N} (\hat{Z}^l - Z^l)^2/N$. Predictions with smaller Brier scores are preferred.

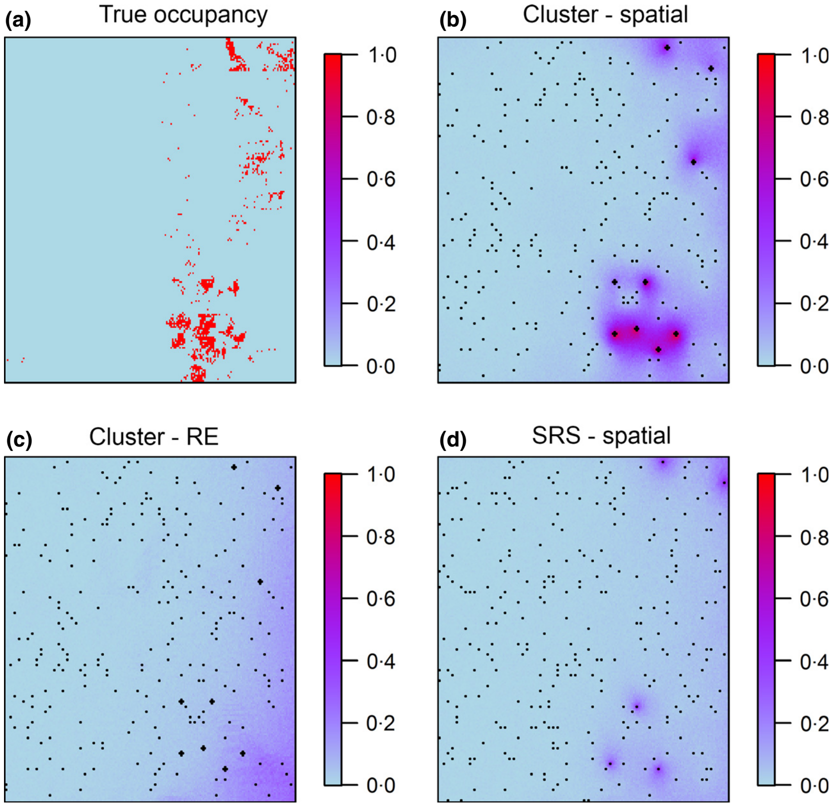

Fig. 1. True occupancy status for the plant *Tamarix ramosissima* from the Inner Mongolia region of China, and posterior mean occupancy probability at non-sampled sites for various designs and analysis methods for one data set (sampling locations are black) with n = 250 and p = 0·25.

Table 4. Bias and MSE for overall occupancy (ψ) and the Brier score for spatial prediction for *Tamarix ramosissima*. Data are sampled using either cluster ('Clu') or simple random sampling ('SRS'), and data are analysed using either spatial ('Sp'), random-effects ('RE') or independent ('In') models with J = 5 sampling occasions. Data generation varies by detection probability (p) and initial sample size (n)

	n	Clu/Sp	Clu/RE	Clu/In	SRS/Sp	SRS/In
BIAS	100	0·006	−0·006	0·032	0·010	0·016
P = 0·25	250	0·007	−0·008	0·036	0·010	0·012
	500	0·002	−0·009	0·040	0·011	0·013
P = 0·75	100	0·000	−0·012	0·026	0·013	0·016
	250	0·001	−0·009	0·041	0·007	0·010
	500	−0·001	−0·008	0·047	0·008	0·010
MSE	100	0·000	0·000	0·003	0·001	0·001
P = 0·25	250	0·000	0·000	0·002	0·000	0·000
	500	0·000	0·000	0·002	0·000	0·000
P = 0·75	100	0·000	0·000	0·002	0·000	0·000
	250	0·000	0·000	0·002	0·000	0·000
	500	0·000	0·000	0·002	0·000	0·000
BR	100	0·028	0·028	0·035	0·030	0·030
P = 0·25	250	0·027	0·028	0·037	0·026	0·028
	500	0·023	0·027	0·036	0·026	0·029
P = 0·75	100	0·028	0·028	0·035	0·028	0·029
	250	0·025	0·028	0·037	0·025	0·028
	500	0·022	0·027	0·036	0·024	0·028

The results (Table 4) mirror the simulation study (Tables 1-3). Fitting the independence model to the clustered data results in a positive bias, and the spatial model with cluster-sampled data has smaller bias and MSE than the spatial model with randomly sampled data. As expected, predictions from the spatial models produce smaller Brier scores than non-spatial models, and on average, the spatial model has smaller Brier score when driven by cluster-sampled data rather than randomly sampled data.

We plotted the posterior distribution of occupancy probabilities $\hat{Z}^l$ ($\Pr(Z = 1|y)$) for one simulated data set with n = 250 and p = 0·25 (Fig. 1, panels 2-4). Comparing the spatial models with cluster and random sampling, the data generated with cluster sampling leads to a more precise distribution map, especially in the south-east, due to the additional samples in occupied regions. In particular, the posterior occupancy probabilities are as high as 0·85 for the cluster-sampled data set compared to a maximum of 0·45 for the randomly sampled data set. This leads to overall improvement in the Brier score. Final sample sizes after adaptive sampling never exceeded 12% additional sampling effort (Table S1).

Discussion

The conservation and management of spatially clustered species is one of the most daunting challenges natural resource managers and ecologists face. It is important that methods are developed that permit reliable inference for these unique scenarios. We have developed an approach that augments the traditional single-season occupancy design to leverage information from adjacent sites when a known detection has occurred. We have provided two models that integrate adaptive cluster sampling into an occupancy estimation framework.

Our simulations show a marked improvement in confidence interval coverage for the new models combined with cluster sampling compared to simple random sampling and traditional occupancy models most notably when detection probability is low and when there is spatial correlation in occupancy. Accounting for the design using the simple cluster random-effects model reduces bias considerably, and full spatial modelling reduces bias further, especially for large n when the spatial covariance parameters can be estimated reliably.

Several other authors have explored the use of different survey designs for occupancy-based studies. MacKenzie & Royle (2005) explored two common sampling designs and their influence on estimator performance. Double sampling, where repeated surveys are conducted at a subset of sites only, was found to have little advantage over the traditional approach while removal sampling, where surveying of a site stops once the species is detected or J surveys have been conducted, was found to be more efficient in terms of obtaining a smaller standard error for estimating occupancy. MacKenzie & Royle (2005) went on further to say that this gain in efficiency for removal sampling was only realized when a greater maximum number of visits to a particular site is conducted. This suggests that the use of these designs is not always warranted except under specific circumstances. We found similar results with our model, and thus, the use of adaptive cluster sampling showed very little improvement in RMSE except when detection probability was low and the initial sample size was low.

Our results suggest that our models may only be useful under certain conditions that relate to specific characteristics of the population. This is not surprising as the benefits of traditional adaptive sampling are only realized for very specific circumstances as well. Several authors have shown that the gain in efficiency for adaptive cluster sampling depends on many factors including the condition to adapt, the number of sites and the aggregation and distribution of the population (Smith, Conroy & Brakhage 1995; Thompson & Seber 1996). Smith, Conroy & Brakhage (1995) showed that for adaptive cluster sampling to be more efficient than simple random sampling the final sample size should not be much larger than the initial sample size. In addition Thompson & Seber (1996) identified a threshold for the initial sample size for which the modified Horvitz-Thompson estimator was more efficient than simple random sampling for binary data ($n > 50$).

Our simulations also support many of the previous findings from the adaptive sampling literature. For example, adaptive cluster sampling may seem daunting for the field biologist when faced with the reality that the sample size is random, which could be a logistical nightmare. This is a common problem for adaptive cluster sampling and has led other authors to develop approaches that provide specific stopping rules or other ways to define a fixed sample size (Christman & Lan 1998, 2001; Rocco 2003). These variations in adaptive sampling may be useful to consider in such cases. This problem is diluted slightly by the argument that sampling adjacent sites in adaptive cluster sampling can be more economical and logistically more feasibly than complete simple random sampling. This has been suggested by other authors (Thompson & Seber 1996).

It is important to note that we do see a small decline in interval coverage with an increase in initial sample size when true occurrence is low. This is most likely due to the fact that ψ corresponds to an estimand of a finite population and the variance of the model-based estimators is not constructed to decline as the proportion of units selected in the sample increases. An alternative approach would be to develop estimators that condition on the event that at least one of the sample units contained the species, as described by Dupuis, Bled & Joachim (2011).

Although Thompson & Seber (1996) found little evidence of improved adaptive sampling estimator performance compared to simple random sampling for binary data, this should not impede the use of such a design with occupancy estimation. The findings of Thompson & Seber (1996) suggest that there is no gain in precision of the adaptive cluster sampling (ACS) estimator, but one benefit is the increase in the number of sites sampled thus increasing the likelihood of sampling more individuals. This gain in the likelihood of observing a particular species has been noted by several other authors as well (Thompson 2004; Salehi & Brown 2010) and can play a critical role in some study objectives when finding a species is equally, if not, more important than estimation. In addition, we found such a stark improvement in interval coverage with minimal variation in model performance over the 500 trials for each simulation it suggests there is a clear advantage in performance compared to simple random sampling.

Here, we have focused solely on occupancy-type data, but we believe that our model can be easily extended. Occupancy estimation has seen many variations as needed to accommodate different objectives and constraints for ecological studies. We believe that many of these same approaches could be easily integrated into our model. For example, the use of auxiliary information collected at each site (e.g. counts of individuals; Royle & Nichols 2003; Royle 2004) could easily be integrated into our model by focusing specifically on abundance instead of occupancy. This would require a different state model in which abundance was directly modelled instead of occupancy or a model that explicitly relied on the occupancy–abundance relationship (Royle & Nichols 2003; Conroy *et al.* 2008). There already exists a wide variety of occupancy-based modelling approaches focused on modelling spatial variation in abundance that could be suggested (Dorazio, Jelks & Jordan 2005; Royle *et al.* 2007; Webster, Pollock & Simons 2008; Post van der Burg *et al.* 2011; Dorazio 2014). As MacKenzie & Royle (2005) found, a removal-based approach may be useful to reduce the logistical effort required to conduct repeat visits while still obtaining reasonable estimates of occupancy. For the cluster design, this may be even more advantageous because sampling adjacent sites and conducting repeat visits can be logistically taxing. Thus, a removal-type design could still provide the benefits of augmenting the design, but would reduce the overall effort.

We envision other areas of expansion that should be investigated as well. As noted in the adaptive sampling literature, changing the definition of the condition to adapt (trigger) can provide valuable changes in estimation and inference (see overview by Turk & Borkowski 2005). We imagine many possible definitions for the trigger in occupancy-based studies which would ultimately depend on the overall objectives of the study. For example, we can conceive of a situation where the use of auxiliary information (e.g. counts of individuals) could be used as the trigger. A second suggestion is to use a combination of species or an index of multiple species (i.e. measure of diversity) as the trigger for adaptation especially if interest is in community composition or species richness. A separate area of expansion involves the exploration of various neighbourhood structures. Christman (1996) found physically contiguous neighbourhoods to be most efficient for classical adaptive sampling and this may be relevant as well. Additionally, we believe these models could be valuable in non-adaptive sampling approaches also. For example, these types of models may work well when there is correlation inherent in the design such as along a transect (similar to Hines *et al*. 2010) and SRS may give misleading results unless the correlated design is incorporated.

Although the incorporation of model-based and design-based approaches is not new, we believe our approach is unique and potentially useful for a variety of studies interested in patchily distributed, clustered or rare species exhibiting spatial variation. This model builds on both the strength of occupancy modelling and adaptive sampling and performs at least as well, and often better, than occupancy modelling alone. In addition, it benefits from incorporating observer behaviour by allowing for extra effort to be included in areas with known detections while permitting statistically rigorous estimates of occupancy and detection probability. We see the continuation of research focusing on integrating sample design and data collection into the modelling framework as a much needed and critical component to rare species conservation and management. Approaches that allow for the flexibility of combining designs and modelling can provide a critical and informative step in conserving and managing rare species.

Acknowledgements

The authors would like to express a special thanks to the Yuancai Lei for providing the data. Additional thanks to the Ministry of Science and Technology and National Natural Sciences Foundation of China for fiscal support for the field work (Project Research Grants 2005DIB5JI42, 31170588 and 30510103195). The comments of the associate editor and two anonymous reviewers substantially improved the manuscript. We would also like to acknowledge the Georgia Cooperative Fish and Wildlife Research Unit for financial assistance. Use of trade, product or firm names does not imply endorsement by the U.S. Government.

Conflict of interest

The authors declare no conflict of interest.

Data accessibility

The data used in this article do not belong to any of the authors of this paper. A request to archive the data was denied by the owner. Authors interested in viewing the data should contact Yuancai Lei (leiycai@yahoo.com; Research Institute of Forest Resource Information Techniques, Chinese Academy of Forestry, Beijing 100091, PR China).

References

Aing, C., Halls, S., Oken, K., Dobrow, R. & Fieberg, J. (2011) A Bayesian hierarchical occupancy model for track surveys conducted in a series of linear, spatially correlated sites. *Journal of Applied Ecology*, **48**, 1508–1517.

Alldredge, M.W., Pollock, K.H., Simons, T.R. & Shriner, S.A. (2007) Multiple-species analysis of point count data: a more parsimonious modeling framework. *Journal of Applied Ecology*, **44**, 281–290.

Bailey, L.L., MacKenzie, D.I. & Nichols, J.D. (2014) Advances and applications of occupancy models. *Methods in Ecology and Evolution*, **5**, 1269–1279.

Brown, J.A., Salehi, M., Moradi, M., Panahbehagh, B. & Smith, D.R. (2013) Adaptive survey designs for sampling rare and clustered populations. *Mathematics and Computers in Simulation*, **93**, 108–116.

Christman, M.C. (1996) Comparison of efficiency of adaptive sampling in some spatially clustered populations. *Proceedings of the Section on Statistics and the Environment*, pp. 122–126. American Statistical Association, Alexandria, Virginia, USA.

Christman, M.C. & Lan, F. (1998) Sequential adaptive sampling designs to estimate abundance in rare populations. *Proceedings of the American Statistical Association, Section on Statistics and Environment*, pp. 87–96.

Christman, M.C. & Lan, F. (2001) Inverse adaptive cluster sampling. *Biometrics*, **57**, 1096–1105.

Conroy, M.J., Runge, J.P., Barker, R.J., Schofield, M.R. & Fonnesbeck, C.J. (2008) Efficient estimation of abundance for patchily distributed populations via two-phase, adaptive sampling. *Ecology*, **89**, 3362–3370.

Cressie, N.A. (1993) *Statistics for Spatial Data*. Wiley, New York.

Cunningham, R.B. & Lindenmayer, D.B. (2005) Modeling count data of rare species: some statistical issues. *Ecology*, **86**, 1135–1142.

Diggle, P.J., Menezes, R. & Su, T.L. (2010) Geostatistical inference under preferential sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **59**, 191–232.

Dorazio, R.M. (2014) Accounting for imperfect detection and survey bias in statistical analysis of presence-only data. *Global Ecology and Biogeography*, **23**, 1472–1484.

Dorazio, R.M., Gotelli, N.J. & Ellison, A.M. (2011) Modern methods of estimating Biodiversity from presence-absence surveys. *Biodiversity Loss in a Changing Planet* (eds O. Grillo & G. Venora), pp. 277–302. InTech, Rijeka, Croatia.

Dorazio, R.M., Jelks, H.L. & Jordan, F. (2005) Improving removal-based estimates of abundance by sampling a population of spatially distinct subpopulations. *Biometrics*, **61**, 1093–1101.

Dorazio, R.M. & Rodríguez, D.T. (2012) A Gibbs sampler for Bayesian analysis of site-occupancy data. *Methods in Ecology and Evolution*, **3**, 1093–1098.

Dorazio, R.M., Royle, J.A., Soderstrom, B. & Glimskar, A. (2006) Estimating species richness and accumulation by modeling species occurrence and detectability. *Ecology*, **87**, 842–854.

Dupuis, J.A., Bled, F. & Joachim, J. (2011) Estimating the occupancy rate of spatially rare or hard to detect species: a conditional approach. *Biometrics*, **67**, 290–298.

Edwards Jr, T.C., Cutler, D.R., Zimmerman, N.E., Geiser, L. & Alegria, J. (2005) Model-based stratifications for enhancing the detection of rare ecological events. *Ecology*, **86**, 1081–1090.

Gardner, B., Reppucci, J., Lucherini, M. & Royle, J.A. (2010) Spatially explicit inference for open populations: estimating demographic parameters from camera-trap studies. *Ecology*, **91**, 3376–3383.

Gelfand, A.E. & Smith, A.F.M. (1990) Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398–409.

Gneiting, T. & Raftery, A.E. (2007) Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, **102**, 359–378.

Guillera-Arroita, G., Morgan, B.J., Ridout, M.S. & Linkie, M. (2011) Species occupancy modeling for detection data collected along a transect. *Journal of Agricultural, Biological, and Environmental Statistics*, **16**, 301–317.

Guillera-Arroita, G., Ridout, M.S., Morgan, B.J. & Linkie, M. (2012) Models for species-detection data collected along transects in the presence of abundance-induced heterogeneity and clustering in the detection process. *Methods in Ecology and Evolution*, **3**, 358–367.

Hines, J.E., Nichols, J.D., Royle, J.A., MacKenzie, D.I., Gopalaswamy, A.M., Kumar, N.S. & Karanth, K.U. (2010) Tigers on trails: occupancy modeling for cluster sampling. *Ecological Applications*, **20**, 1456–1466.

Hoeting, J.A., Leecaster, M. & Bowden, D. (2000) An improved model for spatially correlated binary responses. *Journal of Agricultural, Biological, and Environmental Statistics*, **5**, 102–114.

Hooten, M.B., Larsen, D.R. & Wikle, C.K. (2003) Predicting the spatial distribution of ground flora on large domains using a hierarchical Bayesian model. *Landscape Ecology*, **18**, 487–502.

Johnson, D.S., Laake, J.L. & Ver Hoef, J.M. (2010) A model-based approach for making ecological inference from distance sampling data. *Biometrics*, **66**, 310–318.

Johnson, D.S., Conn, P.B., Hooten, M.B., Ray, J.C. & Pond, B.A. (2013) Spatial occupancy models for large data sets. *Ecology*, **94**, 801–808.

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

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

MacKenzie, D.I., Nichols, J.D., Sutton, N., Kawanishi, K. & Bailey, L.L. (2005) Improving inference in population studies of rare species that are detected imperfectly. *Ecology*, **86**, 1101–1113.

MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L. & Hines, J.E. (2006) *Occupancy Estimation and Modeling*. Academic Press, New York, New York.

Martin, J., Edwards, H.H., Bled, F., Fonnesbeck, C.J., Dupuis, J.A., Gardner, B. *et al.* (2014) Estimating upper bounds for occupancy and number of manatees in areas potentially affected by oil from the Deepwater Horizon oil spill. *PLoS One*, **9**, e91683.

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

Pacifici, K., Dorazio, R.M. & Conroy, M.J. (2012) A two-phase sampling design for increasing detections of rare species in occupancy surveys. *Methods in Ecology and Evolution*, **3**, 721–730.

Pacifici, K., Zipkin, E.F., Collazo, J.A., Irizarry, J.I. & DeWan, A. (2014) Guidelines for *a priori* grouping of species in hierarchical community models. *Ecology and Evolution*, **4**, 877–888.

Pati, D., Reich, B.J. & Dunson, D.B. (2011) Bayesian geostatistical modelling with informative sampling locations. *Biometrika*, **98**, 35–48.

Peyrard, N., Sabbadin, R., Spring, D., Brook, B. & Mac Nally, R. (2013) Model-based adaptive spatial sampling for occurrence map construction. *Statistics and Computing*, **23**, 29–42.

Possingham, H. P., Andelman, S.J., Noon, B.R., Trombulak, S. & Pulliam, H.R. (2001) Making smart conservation decisions. *Conservation Biology: Research Priorities for the Next Decade* (eds M.E. Soule & G.H. Orians), pp. 225–244. Island Press, Washington, D.C.

Post van der Burg, M., Bly, B., Ver Cauteren, T. & Tyre, A.J. (2011) Making better sense of monitoring data from low density species using a spatially explicit modeling approach. *Journal of Applied Ecology*, **48**, 47–55.

R Development Core Team (2014) *R: A language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org.

Rapley, V.E. & Welsh, A.H. (2008) Model-based inferences from adaptive cluster sampling. *Bayesian Analysis*, **3**, 717–736.

Rocco, E. (2003) Constrained inverse adaptive cluster sampling. *Journal of Official Statistics*, **19**, 45–57.

Royle, J.A. (2004) N-mixture models for estimating population size from spatially replicated counts. *Biometrics*, **60**, 108–115.

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

Royle, J.A. & Nichols, J.D. (2003) Estimating abundance from repeated presence-absence data or point counts. *Ecology*, **84**, 777–790.

Royle, J.A., Kery, M., Gautier, R. & Schmid, H. (2007) Hierarchical spatial models of abundance and occurrence from imperfect survey data. *Ecological Monographs*, **77**, 465–481.

Salehi, M. & Brown, J.A. (2010) Complete allocation sampling: an efficient and easily implemented adaptive sampling design. *Population Ecology*, **52**, 451–456.

Smith, D.R., Conroy, M.J. & Brakhage, D.H. (1995) Efficiency of adaptive cluster sampling for estimating density of wintering waterfowl. *Biometrics*, **51**, 777–788.

Smith, D.R., Lei, Y., Walter, C.A., Young, J.A. (2012) Incorporating predicted species distribution in conventional and adaptive sampling designs. *Design and Analysis of Long-term Ecological Monitoring Studies* (eds R.A. Gitzen, J.J. Millspaugh, A.B. Cooper and D.S. Licht), pp. 281–296. Cambridge University Press, New York.

Thompson, S.K. (1990) Adaptive cluster sampling. *Journal of the American Statistical Association*, **85**, 1050–1059.

Thompson, S.K. (2002) *Sampling*. Wiley-Interscience, New York, New York.

Thompson, W.L. (ed.) (2004) *Sampling Rare or Elusive Species*. Island Press, Washington, D. C.

Thompson, S.K. & Seber, G.A.F. (1996) *Adaptive Sampling*. Wiley-Interscience, New York, New York.

Turk, P. & Borkowski, J.J. (2005) A review of adaptive cluster sampling: 1990-2003. *Environmental and Ecological Statistics*, **12**, 55–94.

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

Webster, R.A., Pollock, K.H. & Simons, T.R. (2008) Bayesian spatial modeling of data from avian point count surveys. *Journal of Agricultural, Biological, and Environmental Statistics*, **13**, 121–139.

Zipkin, E.F., DeWan, A. & Royle, J.A. (2009) Impacts of forest fragmentation on species richness: a hierarchical approach to community modeling. *Journal of Applied Ecology*, **46**, 815–822.

Received 29 June 2015; accepted 7 October 2015
Handling Editor: Jana McPherson

Supporting Information

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

Data S1. Computational algorithm.

Data S2. MEE codes.

Table S1. Average final sample size and standard deviation out of 500 simulated data sets for different scenarios.