Hierarchical spatial models for predicting pygmy rabbit distribution and relative abundance

Tammy L. Wilson[1]*, James B. Odei[2], Mevin B. Hooten[2,1] and Thomas C. Edwards Jr[3,1]

[1]*Department of Wildland Resources and Ecology Center, Utah State University;* [2]*Department of Mathematics and Statistics, Utah State University; and* [3]*U.S. Geological Survey Utah Cooperative Fish and Wildlife Research Unit, Department of Wildland Resources, Utah State University, Logan, UT, USA*

Summary

1. Conservationists routinely use species distribution models to plan conservation, restoration and development actions, while ecologists use them to infer process from pattern. These models tend to work well for common or easily observable species, but are of limited utility for rare and cryptic species. This may be because honest accounting of known observation bias and spatial autocorrelation are rarely included, thereby limiting statistical inference of resulting distribution maps.

2. We specified and implemented a spatially explicit Bayesian hierarchical model for a cryptic mammal species (pygmy rabbit *Brachylagus idahoensis*). Our approach used two levels of indirect sign that are naturally hierarchical (burrows and faecal pellets) to build a model that allows for inference on regression coefficients as well as spatially explicit model parameters. We also produced maps of rabbit distribution (occupied burrows) and relative abundance (number of burrows expected to be occupied by pygmy rabbits). The model demonstrated statistically rigorous spatial prediction by including spatial autocorrelation and measurement uncertainty.

3. We demonstrated flexibility of our modelling framework by depicting probabilistic distribution predictions using different assumptions of pygmy rabbit habitat requirements.

4. Spatial representations of the variance of posterior predictive distributions were obtained to evaluate heterogeneity in model fit across the spatial domain. Leave-one-out cross-validation was conducted to evaluate the overall model fit.

5. *Synthesis and applications.* Our method draws on the strengths of previous work, thereby bridging and extending two active areas of ecological research: species distribution models and multi-state occupancy modelling. Our framework can be extended to encompass both larger extents and other species for which direct estimation of abundance is difficult.

Introduction

Ecologists, conservationists and managers often make decisions with incomplete information about the system or species of interest. The predictive spatial distribution model is one tool often used to make decisions based on incomplete information about a species (e.g. Zarnetske, Edwards Jr. & Moisen 2007; La Morgia, Bona & Badino 2008). Such models may be misleading because they can fail to account for biases in species distribution data (Carroll & Johnson 2008) due to imperfect detection (Stauffer, Ralph & Miller 2002) and autocorrelation (Cressie 1993; Hoeting 2009). Recent species distribution models have variously addressed these problems, but only a few have capitalized on count data to produce maps of abundance that also account for both spatial autocorrelation and imperfect detection (e.g. Thogmartin, Sauer & Knutson 2004; Gorresen *et al.* 2009).

Species that are rare or secretive pose a unique set of problems for ecologists interested in using predictive species distribution models. If rarity leads to poor detection, then a high number of zero observations can lead to violations of statistical assumptions of standard generalized linear models (Cunningham & Lindenmayer 2005), and large variances in occupancy models (MacKenzie *et al.* 2009). To improve detection, indirect indices of presence or relative abundance such as burrows, nests, tracks, faecal material or hair samples (signs) may be used (Stanley & Royle 2005). Uncertainty is introduced to the modelling process when indirect detection indices are used to build habitat models, because it is difficult to know if the sign was produced by the organism of interest. For example, a burrow may not be a convincing indicator of presence for one of many burrowing organisms, but when combined with a

*Correspondence author. E-mail: t.w@aggiemail.usu.edu

species-specific observation (like faecal pellets), the burrow may become a more convincing argument for presence. Thus, we observe that all burrows with faecal pellets of any species of interest (burrow utilization) are completely contained within the universe of all burrows on the landscape (burrow intensity), giving these two data levels natural hierarchical structure.

Our objective is to use an empirical Bayesian hierarchical spatial model to produce maps of ecologically important variables: distribution and relative abundance of a small rabbit that is difficult to observe. Our model uses hierarchically related indirect detection data (counts of burrows and presence of faecal pellets) in a joint likelihood that incorporates spatial structure and measurement uncertainty. We show how inference about spatial distribution is enhanced by location-specific uncertainty estimates in unobserved variables (burrow intensity and utilization). This allows us to combine information contained in both levels of data to produce maps of distribution (burrow occupancy) and relative abundance. We then demonstrate the flexibility of our approach by presenting burrow occupancy maps using different assumptions about organismal resource use. While use of hierarchical methods to produce species distribution models is continuing to gain popularity (e.g. Thogmartin, Sauer & Knutson 2004; Gorresen *et al.* 2009), ours builds on previous efforts by using hierarchical indirect detection data as inputs. We also build on multi-level occupancy models (MacKenzie *et al.* 2009) by producing statistically rigorous spatial prediction of distribution and relative abundance.

Materials and methods

STUDY SYSTEM

The pygmy rabbit *Brachylagus idahoensis* Merriam serves as the species of interest in the models we present. Pygmy rabbits occur only in the intermountain western United States, where they depend primarily on big sagebrush *Artemisia tridentata* Nutt. for food and cover. Pygmy rabbits are petitioned to be listed under the United States Endangered Species Act (Fite *et al.* 2003), but little is known about their abundance or distribution. Pygmy rabbits are known to dig their own burrows, and although burrow use is little understood, occupied pygmy rabbit burrows tend to be associated with copious amounts of faecal pellets (Ulmschneider *et al.* 2004). Observing rabbits directly to estimate abundance is difficult because they are small, secretive, difficult to trap and lack markings that would allow for the identification of individuals. As a result, monitoring distribution and relative abundance using indirect indicators of rabbit activity (burrows and faecal pellets) is an attractive alternative to direct observation. Several species distribution models have been proposed for pygmy rabbits using burrows with faecal pellets as inputs (e.g. Simons & Laundré 2004; Himes & Drohan 2007). These previous efforts were successful at delineating important habitat variables, but inference was limited because autocorrelation and detection probability were both ignored in the modelling process.

STUDY AREA

Modelling was conducted at a 21 600-ha site in Rich County, located in northern Utah, USA. The site ranges in elevation from 1800 to 2300 m and is predominately covered by sagebrush-steppe vegetation. Prior to the initiation of sampling, the site was known to be occupied by pygmy rabbits, but their spatial distribution within the study area was largely unknown.

We created a prediction domain consisting of systematically spaced points (300 × 200 m) that formed the centres of 6-ha rectangular polygons for the entire study area. This spacing was based on estimated maximum adult female home-range (Sanchez & Rachlow 2008). The prediction grid cell that overlapped the majority of the area sampled by the burrow transects (described below) was selected as the point for intersecting training data with the grid.

BURROW SAMPLING

We used a modified systematic and stratified design to place 38 sampling locations within the study area. We used a randomly started 2500-m tessellation grid to place 28 of these points. Use of a systematic grid for sampling precludes the estimation of spatial structure that occurred at a finer scale than the tessellation grid (Wintle & Bardos 2006). To combat this problem, we randomly selected an additional 10 sampling locations within soil types that could potentially harbour pygmy rabbits. It should be noted that we sampled a small area of the total domain and prediction would be improved with additional data collection.

We used distance sampling (Buckland *et al.* 2004) to conduct burrow counts at each sampling location along five parallel 200-m line transects spaced 50 m apart (1000 m of total line sampled/site). We randomly selected one of eight bearings (cardinal and inter-cardinal) as the direction the line transects were run at each of the 38 sites. A single observer documented burrows that could be seen directly from the line. Once a burrow was observed, the following were measured by additional observers: (i) distance of the burrow from the line, and (ii) presence or absence of pygmy rabbit faecal pellets. We assumed 100% detection of all burrows directly on the sample transect line.

HIERARCHICAL MODEL

Ecologists build statistical models using data to make inference about a process of interest. In the simplest case, a likelihood-based statistical modelling formulation seeks to maximize the probability of the data given some assumption about the process and parameters. In the case of a generalized linear model, the parameters of the process are the regression coefficients (β). This model assumes that there is only measurement uncertainty, and that it is captured by the error term. If we are aware of uncertainty in a process of interest, then we can use the natural hierarchical structure obtained by factoring the joint distribution of the data and process components to model the uncertainty affecting measurement in a data model separately from that in the process model (Cressie *et al.* 2009). An advantage of this framework for complicated models is that additional data or process levels, and information about their associated uncertainty, can be incorporated easily (Cressie *et al.* 2009).

In this study, we implemented a spatially explicit linear model using a Bayesian hierarchical framework described in detail below. Recall that we are fundamentally interested in making inference about pygmy rabbit activity at a site based on the spatial distribution of two hierarchically related levels of indirect evidence of pygmy rabbit occurrence. The first level of interest is the intensity of all burrows regardless of their use by pygmy rabbits. It was impractical to sample all burrows within our large study area, so we introduced an additional random variable at this level pertaining to the probability of detecting a burrow given it was in the 6-ha

sampling area. The second level of interest is the number of burrows used by pygmy rabbits. We termed this level utilization, where counts of burrows with pygmy rabbit faecal pellets are normalized by the total number of burrows within the domain. We assumed that if present, the probability of detecting burrows with faecal pellets was identical to those without. Utilization was therefore modelled as the probability of finding a burrow with pygmy rabbit faecal pellets given burrows. We present a joint likelihood (or data model) based on our observations of burrow counts and pygmy rabbit use inferred through faecal pellets.

We assumed that both burrow intensity and probability of utilization were related to our set of spatial covariates (Table 1) and potentially subject to correlated spatial error. It is impossible to know *a priori* the nature of the latent spatial structure of either process (intensity or utilization). We therefore used geostatistical methods to estimate spatial dependence parameters in an empirical Bayesian fashion (Casella 1985; see Appendix S1). Such empirical Bayesian approaches have proven useful in cases where fully Bayesian estimation is cumbersome in practice (Hooten, Larsen & Wikle 2003). Specifically, spatial covariance parameters have proven difficult to estimate when deeply nested in hierarchical models and only small amounts of spatial data are available (Carlin & Lewis 2009). In this case, other model parameters can help to absorb any potential uncertainty not accounted for in the empirical Bayes procedure. We also used preliminary analysis (Appendix S1) to obtain the probability of detection for burrows using distance sampling (Buckland *et al.* 2004).

We implemented the hierarchical model using the Bayesian framework described as follows. Let n_i and Y_i represent observed number of burrows with faecal pellets and burrows, respectively, at each location i (where i = 1, 2, …, m). The true number of burrows for each location, N_i, was not observed. We therefore considered the two binomial distributions: $n_i|Y_i,\theta_i \sim \text{Binom}(Y_i,\theta_i)$ and $Y_i|N_i,\varphi \sim \text{Binom}(N_i,\varphi)$, where θ_i is the probability of observing pygmy rabbit faecal pellets given the observed burrows at a specified location i and φ is the probability of burrow detection. Burrow detection is defined as $\Phi^{-1}(\varphi) \sim \text{Norm}(\mu_\varphi,\sigma^2_\varphi)$, as provided by the distance sampling analysis, where $\Phi^{-1}(\varphi)$ denotes the probit transformation of burrow detectability. As part of our process stage in the hierarchical framework, we modelled the true number of burrows, N_i, with a Poisson distribution with intensity λ_i. Noting the availability of our covariates at all locations, we then specify linear models for the log transformation of λ, where $\lambda = (\lambda_1,…, \lambda_m)$, and probit transformation of θ, where $\theta = (\theta_1,…, \theta_m)$. Thus,

$$\log(\lambda) = X_\lambda\beta+\varepsilon \qquad \text{eqn 1}$$

$$\Phi^{-1}(\theta) = X_\theta\alpha + \eta \qquad \text{eqn 2}$$

where, $\mathbf{X}_\lambda$ is $m \times p$, $\mathbf{X}_\theta$ is $m \times q$, β is $p \times 1$, α is $q \times 1$, ε is $m \times 1$ and η is $m \times 1$. The error terms, ε and η were assumed to have multivariate normal distributions $\varepsilon \sim \text{Norm }(\mathbf{0}, \sigma^2_\varepsilon \mathbf{R}_\varepsilon)$ and $\eta \sim \text{Norm }(\mathbf{0}, \sigma^2_\eta \mathbf{R}_\eta)$ where, $\mathbf{R}_\varepsilon$ and $\mathbf{R}_\eta$ are spatial correlation matrices, the forms of which (e.g. exponential, Gaussian or spherical) are dictated by the residual spatial structure in our preliminary data analysis (Appendix S1). The hierarchical model is summarized as follows:

$$n_i|Y_i, \theta_i \sim \text{Binom}(Y_i, \theta_i), i = 1, 2, ..., m \qquad \text{eqn 3}$$

$$Y_i|N_i, \varphi \sim \text{Binom}(N_i, \varphi), i = 1, 2, ..., m \qquad \text{eqn 4}$$

$$N_i|\lambda_i \sim \text{Pois}(\lambda_i), i = 1, 2, ..., m \qquad \text{eqn 5}$$

$$\log(\lambda)|\beta, \sigma^2_\varepsilon \mathbf{R}_\varepsilon \sim \text{Norm}\left(X_\lambda\beta, \sigma^2_\varepsilon \mathbf{R}_\varepsilon\right) \qquad \text{eqn 6}$$

$$\Phi^{-1}(\theta)|\alpha, \sigma^2_\eta \mathbf{R}_\eta \sim \text{Norm}\left(X_\theta\alpha, \sigma^2_\eta \mathbf{R}_\eta\right) \qquad \text{eqn 7}$$

where eqns 3 and 4 form the data model and eqns 5–7 make up the process model. To make our priors on the regression parameters vague, we specified multivariate normal distributions with mean vectors equal to zero and variance components equal to 1000 for β and α. For the variance component σ^2_ε, we specified a conjugate Jeffreys prior as it is difficult to estimate the scale parameter using geostatistics without observing λ. We then obtain the posterior distribution given the number of burrows with faecal pellets (n_i) and burrows (Y_i) as proportional to the product of the likelihood of the data given the latent process models and parameter models shown below.

$$[\{N_i\}, \{\theta_i\}, \{\lambda_i\}, \beta, \alpha, \sigma^2_\varepsilon|\{Y_i\}, \{n_i\}] \propto \prod_{i=1}^m [n_i|Y_i, \theta_i] \times \prod_{i=1}^m [Y_i|N_i, \varphi]$$

$$\times \prod_{i=1}^m [N_i|\lambda_i] \times [\lambda|\beta] \times [\theta|\alpha] \times [\beta] \times [\alpha] \times [\sigma^2_\varepsilon] \qquad \text{eqn 8}$$

The posterior in eqn 8 is not analytically tractable. Thus, given empirical estimates of spatial dependence parameters (υ_ε, υ_η) and burrow detection probability (φ), the model was then implemented using a hybrid Metropolis-Hastings and Gibbs Markov Chain

Table 1. Spatial covariates used in burrow intensity and burrow utilization models. Raster data were summarized for each 6-ha grid cell and the mean value was used for modelling

Variable	Description	Source
SLOPE	Per cent slope	NED
ASPVAL	Aspect transformed to create an index ranging from 0 to 1 where 0 is minimum soil moisture and 1 is maximum soil moisture	NED
ASPWEST	Aspect transformed to create an index ranging from 0 to 1 where 0 is maximum scouring and 1 is maximum deposition	NED
NIR	Near infrared band reflectance	NAIP
RED	Red band reflectance	NAIP
WATER	Euclidian distance to nearest water source (stream or spring)	DLG
X	Easting in Universal Transverse Mercator, zone 12, NAD83	Prediction grid
Y	Northing in Universal Transverse Mercator, zone 12, NAD83	Prediction grid

Data sources include: National elevation data set (NED; 10 m raster), National Agriculture Imagery Program (NAIP; 2 m raster) and Digital Line Graph (vector DLG; ground-truthed with NAIP).

Monte Carlo (MCMC) algorithm using program R (R Development Core Team, 2009). The MCMC algorithm was run for 10 000 iterations after a burn-in period of 2000 iterations. Convergence occurred rapidly and was assessed visually to ensure a stationary posterior distribution was well characterized. MCMC samples from the posterior distribution were used to calculate summary statistics for all latent processes and model parameters. Using composition sampling, posterior predictive distributions were obtained for the latent parameters θ (burrow utilization), and λ (burrow intensity) for the entire spatial domain. The details of the MCMC algorithm are beyond the scope of this paper, and interested readers are referred one of many texts on the subject (e.g. Banerjee, Carlin & Gelfand 2003; Carlin & Lewis 2009).

MODEL VALIDATION

We evaluated the spatial precision of the posterior expectations of our latent processes (burrow intensity and burrow utilization) by finding the standard deviations of the predicted posterior distributions. We mapped the standard deviations to assess the heterogeneity of predictive precision throughout the site. Maps of posterior predictive standard deviations do not evaluate overall model performance. We therefore performed a leave-one-out cross-validation procedure to formally evaluate overall model fit. We omitted the data for one of the 38 sites, and then re-ran the model obtaining posterior predictions of the expected data (n_i and Y_i) for the omitted site. It was only possible to evaluate the counts (n_i and Y_i) using cross validation because all other processes were unobserved.

Results

The predictive map of burrow intensity (λ) shows high burrow intensity in the valley bottoms of the study area (Fig. 1a). Recall that burrow intensity represents all burrows, regardless of origin, expected to occur within each 6-ha grid cell. High burrow intensity was predicted for slopes near drainages of perennial streams within the study area. Covariate relationships were considered statistically significant if 95% credible intervals did not overlap zero (Table 2). Burrow intensity was positively associated with northing (Y), slope (SLOPE), and soil moisture (ASPVAL), and negatively associated with easting (X), near infrared reflectance (NIR), and distance to water (WATER) (Table 2).

We modelled utilization (θ) as the proportion of the burrows that were expected to contain evidence of pygmy rabbit. High utilization was observed along the slopes and ridge tops within the central portion of the spatial domain (Fig. 1b). Given burrows, utilization was positively related to soil and snow deposition (ASPWEST) and red band reflectance (RED), and negatively related to easting (X), near infrared reflectance (NIR), and distance to water (WATER) (Table 2). Slope (SLOPE) and soil moisture (ASPVAL) were not shown to be statistically significant because regression coefficient 95% credible intervals overlapped zero.

While burrow utilization (θ) and intensity (λ) were processes leading to our observations, those processes can also be used to learn about the total number of burrows expected to have pygmy rabbit faecal pellets. If we let Z represent the true number of burrows with faecal pellets, then conditioned on the true number of burrows (N_i) and probability of utilization (θ_i), Z_i comes from a binomial distribution: $Z_i \sim \text{Binom}(N_i, \theta_i)$. We can thus obtain predictions for pygmy rabbit burrow abundance over the entire spatial domain. If we assume that the pygmy rabbit burrow abundance (Z) is proportional to the number of pygmy rabbits, then Z becomes a metric of relative pygmy rabbit abundance. Maps of relative abundance (Fig. 1c) shows the expectation of pygmy rabbit burrow abundance attributable to spatial covariates and spatial autocorrelation formally included in our model. Pygmy rabbit burrow abundance is predicted to be higher in both valleys and slopes in the centre of the domain. An unsampled region in the western edge of the domain is also expected to have high pygmy rabbit abundance.

We gain insight about pygmy rabbit occupancy within the spatial domain if we assume that a site with at least one burrow with pygmy rabbit faecal pellets is occupied by pygmy rabbits. Thus, we compute the posterior probability that a site has at least one burrow with evidence of pygmy rabbits given expected count data: $p(Z > 0|\mathbf{n}, Y)$. Maps of pygmy rabbit burrow occupancy (Fig. 2a) show that burrows with evidence of pygmy rabbits are expected to be found throughout the spatial domain. However, there are areas where the probability of pygmy rabbit burrow occupancy is expected to be reduced, indicating within site heterogeneity in pygmy rabbit distribution of a spatial domain that is known to be occupied. In particular, the agricultural fields on the east side of the domain are expected to have low probabilities of pygmy rabbit occupancy.

We also demonstrated how different assumptions about the number of burrows required to indicate animal presence affect the maps of pygmy rabbit distribution. If the number of burrows with faecal pellets required to indicate pygmy rabbit presence was changed from one to five, $p(Z \geq 5|\mathbf{n}, Y)$, then the total area expected to be used by pygmy rabbits was reduced (Fig. 2b). If this occupancy criterion was further increased to 10, $p(Z \geq 10|\mathbf{n}, Y)$, the total area expected to be occupied was again reduced, and the number of occupied polygons connected by $P > 0.5$ was changed from one to two (Fig. 2c).

Maps of standard deviations were produced from the marginal predictive distributions for the intensity and utilization processes (Fig. 3). As expected, maps of burrow intensity show that prediction precision was best in regions of the map where data collection points were close together, and less precise in areas of the map with sparse data and in areas of extrapolation. Recall that burrow intensity (λ) was modelled with a log-linear model, where the mean and variance are expected to have a one-to-one relationship as an artefact of the Poisson distribution. This can be seen in the similarity of appearance of the prediction maps of standard deviation (Fig. 3) and mean (Fig. 1). The appearance of the mean (Fig. 1) and standard deviation (Fig. 3) prediction maps for the utilization parameter (θ) are also quite similar. This is because the maximum proportion of all burrows that are expected to have pygmy rabbit faecal pellets was just over 50%. The variance of the binomial distribution is maximum at $\theta = 0.50$ and falls as the probability of the event of interest occurring (or not) is more certain. Leave-one-out cross validation revealed that predictions for burrows and burrows with faecal pellets missed the observed values at
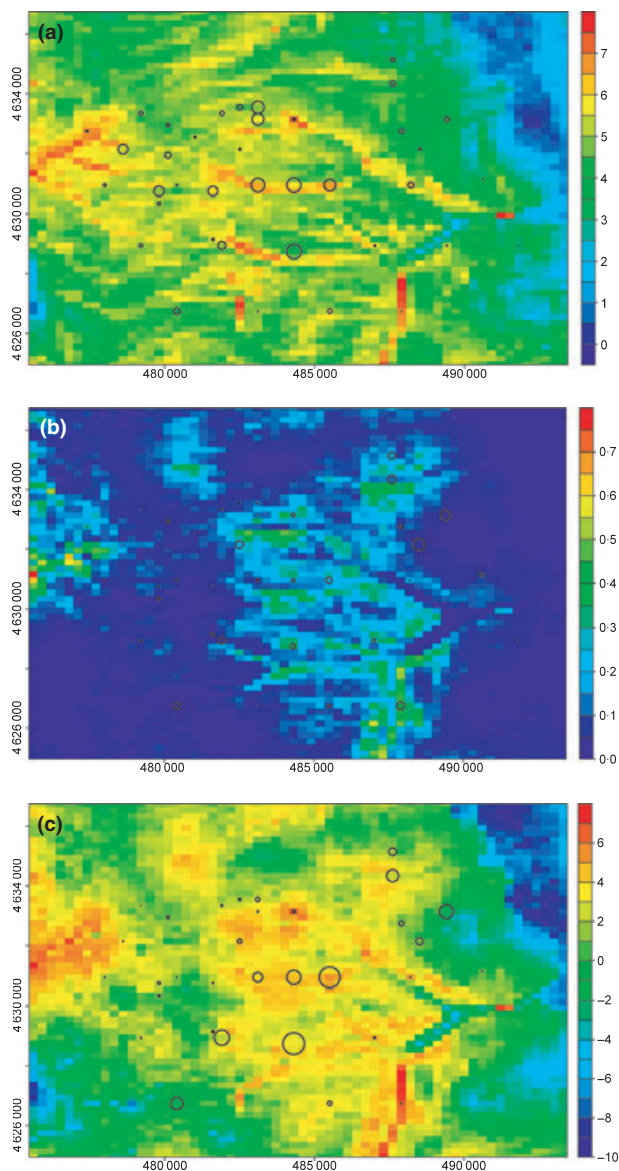
Fig. 1. Map (a) depicts the predicted mean of the natural logarithm of burrow intensity (λ). Map (b) depicts the predicted mean of burrow utilization (θ). Map (c) depicts the predicted mean number of pygmy rabbit burrows (presented on the log scale- negative numbers indicate mean values < 1) an index of pygmy rabbit burrow abundance. Coordinates listed on the margin of the map indicate the boundaries of the study domain and are projected: UTM (Zone 12, NAD 83). The colours on the map present a gradient from cool to warm representing the modelled expected value of the parameter in each grid cell from low values to high. Circles within each map represent the value of each parameter calculated directly using field data. The size of the circle corresponds to the estimated value of the parameter.

some of the sites. However, the predictions followed the general pattern of the omitted observations (solid line in Fig. 4), suggesting that the model captured the overall behaviour well.

Discussion

We used our hierarchical model to successfully incorporate two sources of indirect data in a coupled likelihood and create posterior predictive distributions for the processes of interest
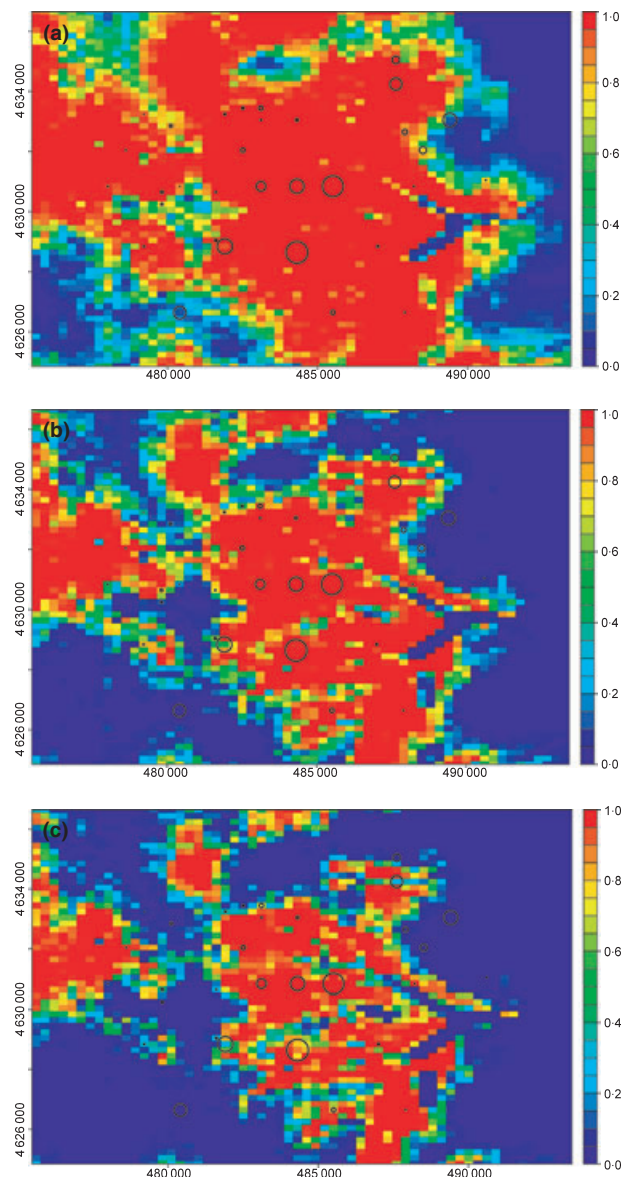
Fig. 2. Maps of pygmy rabbit occupancy using three criteria for assuming a site is occupied. Map (a) depicts the probability of a single burrow with pygmy rabbit faecal pellets given predicted counts of burrows and utilization $p(\mathbf{Z} > 0|\mathbf{n}, \mathbf{Y})$. Map (b) depicts the probability of 5 burrows $p(\mathbf{Z} \geq 5|\mathbf{n}, \mathbf{Y})$, and map (c) depicts the probability of 10 burrows $p(\mathbf{Z} \geq 10|\mathbf{n}, \mathbf{Y})$. Coordinates listed on the margin of the map indicate the boundaries of the study domain and are projected: UTM (Zone 12, NAD 83). The colours on the map present a gradient from cool to warm representing the modelled probability of the parameter in each grid cell from low values to high. Circles within each map represent the value of each parameter calculated directly using field data. The size of the circle corresponds to the estimated value of the parameter.

over the spatial domain. This framework allowed us to present maps of ecologically important variables such as relative abundance and occupancy. Our model is similar to previous hierarchical Bayesian species distribution models that incorporate spatial structure (e.g. Hooten *et al.* 2003; Latimer *et al.* 2006; Carroll & Johnson 2008; Howell, Peterson & Conroy 2008), except that, these used presence–absence data and were limited

Table 2. Regression coefficients (β, α) for the burrow intensity [log(λ)] and pygmy rabbit utilization [$\Phi^{-1}(\theta)$]

	Mean	Standard deviation	Lower 95% credible interval	Upper 95% Credible Interval
Burrow intensity log(λ)				
X	$-1\cdot52 \times 10^{-04}$	$2\cdot92 \times 10^{-06}$	$-1\cdot57 \times 10^{-04}$	$-1\cdot46 \times 10^{-04}$
Y	$6\cdot64 \times 10^{-05}$	$4\cdot52 \times 10^{-06}$	$5\cdot77 \times 10^{-05}$	$7\cdot54 \times 10^{-05}$
SLOPE	$6\cdot12 \times 10^{-02}$	$2\cdot20 \times 10^{-03}$	$5\cdot69 \times 10^{-02}$	$6\cdot56 \times 10^{-02}$
ASPVAL	$6\cdot02 \times 10^{-01}$	$3\cdot52 \times 10^{-02}$	$4\cdot33 \times 10^{-01}$	$6\cdot70 \times 10^{-01}$
NIR	$-4\cdot22 \times 10^{-02}$	$1\cdot23 \times 10^{-03}$	$-4\cdot46 \times 10^{-02}$	$-3\cdot97 \times 10^{-02}$
WATER	$-6\cdot89 \times 10^{-04}$	$1\cdot81 \times 10^{-05}$	$-7\cdot25 \times 10^{-04}$	$-6\cdot56 \times 10^{-04}$
Utilization $\Phi^{-1}(\theta)$				
X	$-1\cdot66 \times 10^{-04}$	$5\cdot17 \times 10^{-05}$	$-2\cdot67 \times 10^{-04}$	$-6\cdot66 \times 10^{-05}$
SLOPE	$3\cdot02 \times 10^{-02}$	$2\cdot54 \times 10^{-02}$	$-1\cdot93 \times 10^{-02}$	$8\cdot05 \times 10^{-02}$
ASPVAL	$-7\cdot14 \times 10^{-01}$	$4\cdot92 \times 10^{-01}$	$-1\cdot68 \times 10^{-00}$	$2\cdot38 \times 10^{-01}$
ASPWEST	$1\cdot74 \times 10^{-00}$	$7\cdot75 \times 10^{-01}$	$2\cdot16 \times 10^{-01}$	$3\cdot38 \times 10^{-00}$
NIR	$-7\cdot35 \times 10^{-01}$	$2\cdot85 \times 10^{-02}$	$-1\cdot30 \times 10^{-01}$	$-1\cdot95 \times 10^{-02}$
RED	$6\cdot39 \times 10^{-02}$	$1\cdot70 \times 10^{-02}$	$3\cdot13 \times 10^{-02}$	$9\cdot78 \times 10^{-02}$
WATER	$-6\cdot77 \times 10^{-04}$	$1\cdot80 \times 10^{-04}$	$-1\cdot03 \times 10^{-03}$	$-3\cdot27 \times 10^{-04}$

Regression coefficients where the 95% credible interval does not overlap 0 are highlighted with bold text.
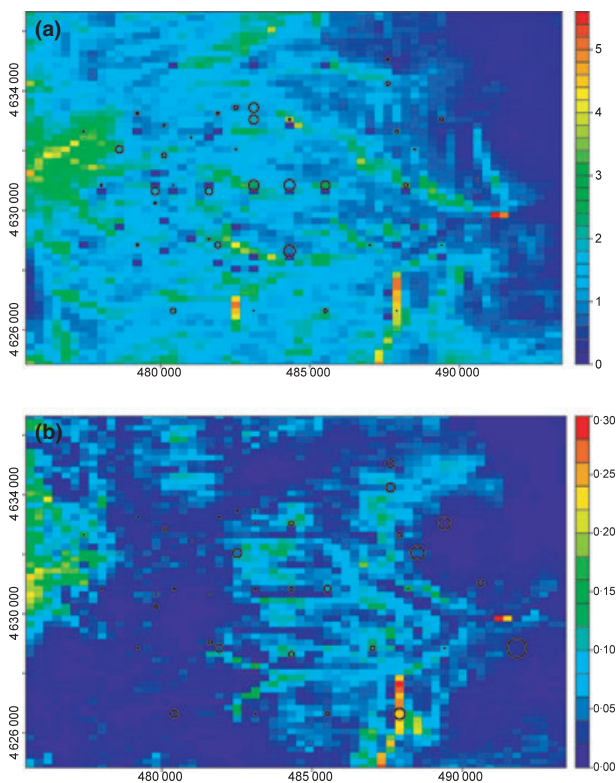

Fig. 3. Standard deviation of predictive distributions for burrow intensity (a; presented on the log scale) and burrow utilization (b). The colours on the map present a gradient from cool to warm representing the modelled standard deviation of the parameter in each grid cell from low values to high. Circles within each map represent the value of each parameter calculated directly using field data. The size of the circle corresponds to the estimated standard deviation of the parameter.

tial structure (e.g. Thogmartin, Sauer & Knutson 2004; Gorresen *et al.* 2009). Our approach is similar to the above studies, except that we use counts of indirect detection indices and a geostatistical approach to estimate the spatial covariance structure. The use of geostatistics rather than Markov random field models allowed us to treat the landscape as a continuous gradient of habitat, rather than a series of classified habitat polygons, which may be conceptually more desirable (Manning, Lindenmayer & Nix 2004). The resulting posterior predictive distribution maps account for both process and observation uncertainty, and provide a framework for improving ecological inference while informing conservation and management decisions.

Our approach also draws on the strengths of multi-state occupancy models (e.g. Nichols *et al.* 2007; MacKenzie *et al.* 2009), but with two important distinctions. First, our model uses signs that persist in the environment, and we can not assume that burrows with faecal pellets are currently occupied. However, a burrow with faecal pellets does indicate use by our species of interest at some point in the recent past, and can safely be considered important for habitat modelling. Secondly, our model explicitly accounts for spatial autocorrelation, where multi-state occupancy models as presented in MacKenzie *et al.* (2009) do not. This allowed us to use geostatistics to produce statistically rigorous prediction maps of our unobserved processes of interest. While MacKenzie *et al.* (2009) argue that careful study design and the use of model averaging precludes the need to account for spatial structure in occupancy models, our approach follows that of Hoeting (2009) who argues that even in well-designed studies of a spatial process, misspecification is likely if autocorrelation is ignored during model selection. Our model therefore provides a useful extension of multi-state occupancy models by formally accounting for spatial structure.

The relative abundance map presented in Fig. 1c takes advantage of the number of burrows estimated using the intensity parameter, and the number expected to have sign of
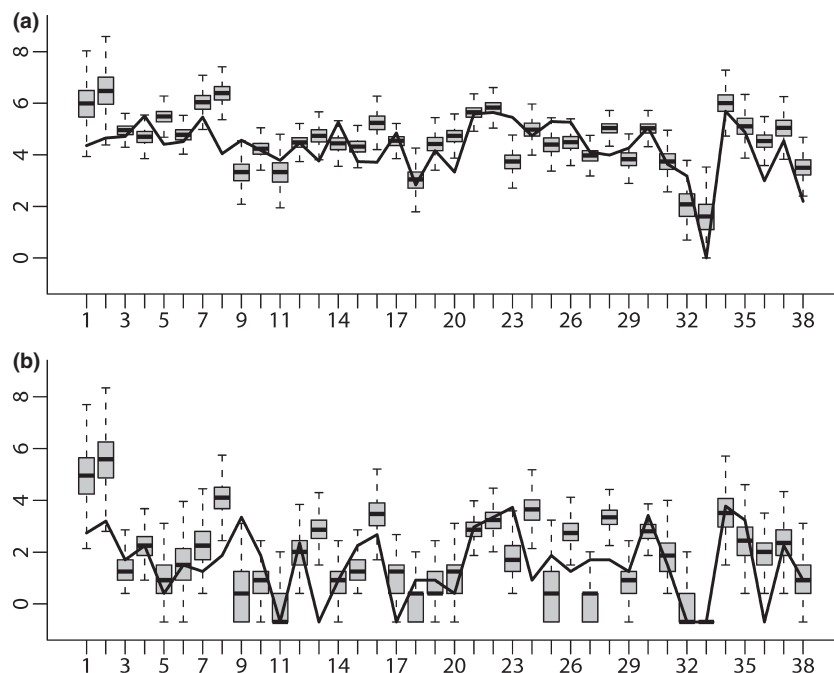
to making inferences on distribution. The use of count data allows for the estimation of relative abundance represented over space while accounting for both observation bias and spa-

Fig. 4. Leave-one-out cross validation results for the observed counts of burrows (a) and burrows with sign (b). The boxplots indicate the median, first and third quartiles, and a rough estimate of the 95% credible intervals of the posterior predictive distribution. The solid line connects the actual observations.

pygmy rabbit activity as estimated by the utilization parameter. Although a similar, but *ad hoc*, metric would have been easily calculated by multiplying the burrow density obtained from program DISTANCE by the ratio of the number of burrows observed to have faecal pellets, our metric provides a statistically rigorous estimate of relative abundance over the entire spatial domain. The autecology of the species of interest could then be used to link relative abundance to actual abundance. For example, if the number of burrows used by a single pygmy rabbit were known, then we could use that information to produce maps of actual abundance.

Our methods produced probabilistic maps pygmy rabbit burrow distribution (Fig. 2a). This map is based on the assumption that a single burrow showing faecal pellets is occupied. This assumption may not be valid if pygmy rabbits use more than one burrow (Sanchez & Rachlow 2008). We therefore demonstrate the flexibility of our approach by presenting two additional maps (Fig. 2b,c) of pygmy rabbit occupancy using different criteria for determining site occupancy. We do not know which of the maps are 'correct' because the minimum number of burrows showing sign in a currently occupied pygmy rabbit home range is not known. However, we demonstrate how the autecology of the focal species could be used to inform the choice of an occupancy criterion used for determining presence in the distribution maps.

We were primarily interested in optimal spatial prediction, but examination of covariates gives some information about the habitat variables important to burrowing animals and pygmy rabbits within our study domain. Inferences about habitat variables made from our model are specific to our study domain, although comparison with previous studies may highlight useful general habitat requirements. In addition, inferences about pygmy rabbit/habitat relationships are limited to the spatial covariates that were available for each grid location within the spatial domain. Therefore, direct resource selection

of variables that may relate to fitness (food and cover) cannot be made at this level. We therefore use surrogates such as aspect, distance to water and spectral reflectance, which may not be directly linked to the biology of the organism in question. A benefit of our approach is that we are able to evaluate the factors influencing burrow intensity separately from those affecting use by pygmy rabbits. These factors are likely to be confounded in studies which model burrows and scat simultaneously.

In the case of our study, interpretation of the regression parameters reveals that both burrow intensity and utilization were negatively associated with near infrared reflectance and distance to water. This indicates that burrowing animals in general and pygmy rabbits specifically select habitat near perennial water sources, but do not occur in riparian habitat or in agricultural fields (types that absorb near infrared radiation). Burrow intensity was also positively associated with slope and aspects related to soil moisture. Given the presence of burrows, pygmy rabbits were positively associated with increased red reflectance, which is consistent with their reliance on sagebrush (sagebrush is the least red-band absorbent of all major plants present at the site). Further they were associated with potential snow and soil deposition (easterly aspects), possibly reflecting microsite preferences of pygmy rabbits that are separate from burrowing animals in general. Previous models using untransformed aspect values show inconsistent results, prompting Rachlow & Svancara (2006) to recommend that aspect not be used in large-scale predictive models of pygmy rabbits. This lack of consistency could indeed be related to differences between study areas, or could illustrate confusion resulting from the confounding of burrow intensity and utilization of traditional species distribution models.

Evaluation of model performance is a vital part of any modelling effort. The many methods available for the evaluation of spatial distribution model performance include

a number of cross-validation techniques. However, model fit within the prediction surface is assumed to be homogenous for all traditional model-fit metrics. A benefit of the Bayesian approach is that inference is made based on the posterior distribution, and the standard deviation of the posterior distribution can be used to assess spatial precision of modelled expectations. Evaluating the map of standard deviations can be used to construct optimal sampling designs for spatio-temporal monitoring (Hooten *et al.* 2009). For example, our processes of interest were modelled from the Poisson distribution where variance has a one-to-one relationship with the mean. An optimal design could be one where sites with expected higher abundance were sampled more intensively than those with lower expected abundance.

SYNTHESIS AND APPLICATION

The modelling framework we presented here is applicable to any species for which multiple levels of indirect detection are available. For example, models of cavity-nesting birds based on snag-density (e.g. Ohmann, McComb & Zumrawi 1994) or actual cavities (e.g. Lawler & Edwards Jr. 2002) can benefit from this approach. Our method is also scaleable to different spatial grains and extents through collection of additional data.

For species where indirect observations have natural hierarchical structure and where direct species detection is low, our method is attractive for surveying large spatial extents. This is because variance of occupancy model parameters increases as detection decreases (MacKenzie & Royle 2005). To combat this problem MacKenzie & Royle (2005) suggest adding additional visits to a site before increasing the number of sites sampled, thereby increasing the precision of occupancy estimates. If the goal of the study is to make spatial prediction of species occurrences, then there is a trade-off between increasing detection, which increases precision, and increasing the number of sites, which improves prediction. Our method allows us to estimate both distribution and relative abundance efficiently while still accounting for imperfect detection, thereby bridging and extending two active areas of ecological research: species distribution models and occupancy modelling.

Acknowledgements

References

Banerjee, S., Carlin, B.P. & Gelfand, A.E. (2003) *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC, Boca Raton.

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L. & Thomas, L. (2004) *Advanced Distance Sampling: Estimating Abundance of Biological Populations*. Oxford University Press, Oxford.

Carlin, B.P. & Lewis, T.A. (2009) *Bayesian Methods for Data Analysis*, 3rd edn. Chapman & Hall/CRC, Boca Raton.

Carroll, C. & Johnson, D.S. (2008) The importance of being spatial (and reserved): assessing Northern Spotted Owl habitat relationships with hierarchical Bayesian models. *Conservation Biology*, **22**, 1026–1036.

Casella, G. (1985) An introduction to empirical Bayes data analysis. *American Statistician*, **39**, 83–87.

Cressie, N. (1993) *Statistics for Spatial Data*, revised edn. Wiley-Interscience, New York.

Cressie, N., Calder, C.A., Clark, J.S., Ver Hoef, J.M. & Wikle, C.K. (2009) Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. *Ecological Applications*, **19**, 553–570.

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

Fite, K.J.M., Salvo, M., Marlette, B., Kessler, J., Smith, J. & Criddle, C. (2003) *A Petition for Rules to List the Pygmy Rabbit (Brachylagus idahoensis) Occurring in the Conterminous Intermountain and Great Basin Region as Threatened or Endangered Under the Endangered Species Act 16 USC 1531 seq. U.S. Fish and Wildlife Service*. Committee for the High Desert, Boise.

Gorresen, P.M., McMillan, G.P., Camp, R.J. & Pratt, T.K. (2009) A spatial model of bird abundance as adjusted for detection probability. *Ecography*, **32**, 291–298.

Himes, J.G. & Drohan, P.J. (2007) Distribution and habitat selection of the pygmy rabbit, *Brachylagus idahoensis*, in Nevada (USA). *Journal of Arid Environments*, **68**, 371–382.

Hoeting, J.A. (2009) The importance of accounting for spatial and temporal correlation in analyses of ecological data. *Ecological Applications*, **19**, 574–577.

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.

Hooten, M.B., Wikle, C.K., Sheriff, S.L. & Rushin, J.W. (2009) Optimal spatio-temporal hybrid sampling designs for ecological monitoring. *Journal of Vegetation Science*, **20**, 639–649.

Howell, J.E., Peterson, J.T. & Conroy, M.J. (2008) Building hierarchical models of avian distributions for the State of Georgia. *Journal of Wildlife Management*, **72**, 168–178.

La Morgia, V., Bona, F. & Badino, G. (2008) Bayesian modelling procedures for the evaluation of changes in wildlife habitat suitability: a case study of roe deer in the Italian Alps. *Journal of Applied Ecology*, **45**, 863–872.

Latimer, A.M., Wu, S., Gelfand, A.E. & Silander Jr, J.A. (2006) Building statistical models to analyze species distributions. *Ecological Applications*, **16**, 33–50.

Lawler, J.J. & Edwards Jr, T.C. (2002) Landscape patterns as habitat predictors: building and testing models for cavity-nesting birds in the Uinta Mountains of Utah, USA. *Landscape Ecology*, **17**, 233–245.

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., Seamans, M.E. & Gutiérrez, R.J. (2009) Modeling species occurrence dynamics with multiple states and imperfect detection. *Ecology*, **90**, 823–835.

Manning, A.D., Lindenmayer, D.B. & Nix, H.A. (2004) Continua and umwelt: novel perspectives on viewing landscapes. *Oikos*, **104**, 621–628.

Nichols, J.D., Hines, J.E., MacKenzie, D.I., Seamans, M.E. & Gutiérrez, R.J. (2007) Occupancy estimation with multiple states and state uncertainty. *Ecology*, **88**, 1395–1400.

Ohmann, J.L., McComb, W.C. & Zumrawi, A.A. (1994) Snag abundance for primary cavity-nesting birds on nonfederal forest lands in Oregon and Washington. *Wildlife Society Bulletin*, **22**, 607–620.

R Development Core Team. (2009) *R: A Language and Environment for Statistical Computing*, 2.9.1 edn. R Foundation for Statistical Computing, Vienna. Available at: http://www.R-project.org (accessed on 15 April 2009).

Rachlow, J.L. & Svancara, L.K. (2006) Prioritizing habitat for surveys of an uncommon mammal: a modeling approach applied to pygmy rabbits. *Journal of Mammalogy*, **87**, 827–833.

Sanchez, D.M. & Rachlow, J.L. (2008) Spatio-temporal factors shaping diurnal space use by pygmy rabbits. *Journal of Wildlife Management*, **72**, 1304–1310.

Simons, E.M. & Laundré, J.W. (2004) A large-scale regional assessment of suitable habitat for pygmy rabbits in southwestern Idaho. *Northwest Science*, **78**, 33–41.

Stanley, T.R. & Royle, J.A. (2005) Estimating site occupancy and abundance using indirect detection indices. *Journal of Wildlife Management*, **69**, 874–883.

Stauffer, H.B., Ralph, C.J. & Miller, S.L. (2002) Incorporating detection uncertainty into presence-absence surveys for marbled murrelet. *Predicting Species Occurrences: Issues of Accuracy and Scale* (eds J.M. Scott, P.J. Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall & F.B. Samson), pp. 357–365. Island Press, Washington.

Thogmartin, W.E., Sauer, J.R. & Knutson, M.G. (2004) A hierarchical spatial model of avian abundance with application to Cerulean Warblers. *Ecological Applications*, **14**, 1766–1779.

Ulmschneider, H., Hayes, D., Roberts, H.B., Rachlow, J.L., Forbes, T., Himes, J.G., Sequin, E., Haworth, M., Katzner, T.E., Kozlowski, A., Rauscher, R.L. & Lauridson, P. (2004) *Surveying for Pygmy Rabbits (Brachylagus idahoensis), Fourth Draft*. Bureau of Land Management, Boise.

Wintle, B.A. & Bardos, D.C. (2006) Modeling species–habitat relationships with spatially autocorrelated observation data. *Ecological Applications*, **16**, 1945–1958.

Zarnetske, P.L., Edwards Jr, T.C. & Moisen, G. (2007) Habitat classification modeling with incomplete data: pushing the habitat envelope. *Ecological Applications*, **17**, 1714–1726.

Supporting information

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

Appendix S1. Preliminary data analysis for empirical Bayesian hierarchical model.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.