Bayesian Methods for Hierarchical Distance Sampling Models

C.S. OEDEKOVEN, S.T. BUCKLAND, M.L. MACKENZIE, R. KING, K.O. EVANS, and L.W. BURGER Jr.

The few distance sampling studies that use Bayesian methods typically consider only line transect sampling with a half-normal detection function. We present a Bayesian approach to analyse distance sampling data applicable to line and point transects, exact and interval distance data and any detection function possibly including covariates affecting detection probabilities. We use an integrated likelihood which combines the detection and density models. For the latter, densities are related to covariates in a log-linear mixed effect Poisson model which accommodates correlated counts. We use a Metropolis-Hastings algorithm for updating parameters and a reversible jump algorithm to include model selection for both the detection function and density models. The approach is applied to a large-scale experimental design study of northern bobwhite coveys where the interest was to assess the effect of establishing herbaceous buffers around agricultural fields in several states in the US on bird densities. Results were compared with those from an existing maximum likelihood approach that analyses the detection and density models in two stages. Both methods revealed an increase of covey densities on buffered fields. Our approach gave estimates with higher precision even though it does not condition on a known detection function for the density model.

Key Words: Designed experiments; Hazard-rate detection function; Heterogeneity in detection probabilities; Metropolis–Hastings update; Point transect sampling; RJMCMC.

1. INTRODUCTION

Bayesian methods are becoming increasingly popular for modelling wildlife populations and abundances (e.g. Buckland, Goudie, and Borchers 2000, Marcot et al. 2001; Durban and Elston 2005; Schmidt et al. 2009; King et al. 2010). However, few distance

C.S. Oedekoven (✉) is a Research Fellow in Statistics (E-mail: *cso2@st-andrews.ac.uk*), S.T. Buckland is a Professor of Statistics, M.L. Mackenzie is Lecturer in Statistics, and R. King is a Reader in Statistics, Centre for Research into Ecological and Environmental Modelling, School of Mathematics and Statistics, University of St Andrews, St Andrews, KY16 9LZ, UK. K.O. Evans is a former Research Associate and L.W. Burger Jr. is a Professor of Wildlife Ecology and Management, Department of Wildlife, Fisheries & Aquaculture, Mississippi State University, Box 9690, Mississippi State, MS 39762, USA.

sampling studies have taken a Bayesian approach. Karunamuni and Quinn (1995) developed a Bayes estimator for $f(0)$ using a half-normal detection function (and a gamma prior), where $f(0)$ is a quantity estimated from the distance data that allows observed counts to be adjusted for imperfect detection. The approach makes use of the conjugate property between the normal and gamma distributions. Other studies have built upon this approach. Eguchi and Gerrodette (2009) extended this model by including a binomial likelihood for the encounter rate along the line and described a joint posterior distribution for the density model and effective strip width. Gimenez et al. (2009) implemented an estimator for $f(0)$ using BUGS software, while Zhang (2011) developed an empirical Bayes estimator for $f(0)$. These studies follow a similar approach in that they present their methods for line-transect data and use the half-normal detection function. We describe a Bayesian approach to density estimation from distance sampling data for both line and point transect data applicable to any detection function that uses a hierarchical modelling approach.

Hierarchical distance sampling models have also been developed by e.g. Royle and Dorazio (2008, Chapter 7.1). These authors employ a likelihood that combines the detection function and the binomial model using the marginal probability of encounter (estimated from the detection function) and a data augmentation approach for unobserved groups. The data augmentation approach was adopted by Schmidt et al. (2012), who added a group size model, and Conn, Laake, and Johnson (2012), who extended the approach for double observer data.

In contrast, we use an integrated likelihood that combines the likelihood components of the detection and count models. For the latter, we use a Poisson likelihood for the distance sampling counts that incorporates a component corresponding to the detection function, thus allowing for undetected animals on the surveyed strip (line transects) or circular plot (point transects). In comparison to Eguchi and Gerrodette (2009), who use a binomial likelihood to scale up from density at the line to density in the study area, our Poisson model relates animal counts to covariates via a log-link function. This approach does not rely on random placement of samplers in the study area to the same extent as the design-based approach for the binomial model (Hedley and Buckland 2004). Similar to Chelgren et al. (2011), we include a random effect for site in the Poisson model to accommodate correlated counts due to e.g. repeat counts at the same site. The parameter space is explored using a Metropolis–Hastings (MH) updating algorithm so that different prior distributions for the parameters are easily implemented, and a reversible jump Markov chain Monte Carlo (RJMCMC) algorithm allows for model uncertainty to be incorporated. This may include different key functions for the detection function model and different covariate combinations for both the detection function and the count models.

These developments were motivated by a large-scale experimental study to assess the effects of establishing conservation buffers along field margins on density of several species of conservation interest such as the northern bobwhite (*Colinus virginianus*). Pairs of points were set up at the edge of fields in farmland in 13 states in the USA. These pairs of points consisted of one point on a buffered treatment field and one on a nearby non-buffered control field and will be referred to as sites in the following. Point transect surveys of coveys (fall–winter stable social units of 10–15 individual birds) were conducted at least once but up to three times per year in autumn from 2006 to 2008.

In the following we begin by developing the integrated likelihood (Section 2), describe the Bayesian approach (Section 3) and analyse bobwhite covey data using our Bayesian approach (and a maximum likelihood approach for comparison) (Section 4). Lastly, we contrast our Bayesian approach with existing studies using distance sampling likelihoods (Section 5).

2. AN INTEGRATED LIKELIHOOD FOR DISTANCE SAMPLING DATA

To obtain abundance estimates of a population of interest using distance sampling methods, lines or points may be placed in the study area according to some design (see Buckland et al. 2001, for details). Each line or point is surveyed at least once following the distance sampling protocol where the observer travels along the line (line transects) or remains at the point for a fixed amount of time (point transects). Detections are recorded along with the perpendicular distance from the line to the detection or radial distance from the point to the detection. These distances may be recorded exactly or in predetermined distance intervals. Thus, surveys of this type produce two types of data: firstly, the observed distances y_e with $e = 1, 2, 3, \ldots, n$ (n being the total number of detections) or observed distances n_i in each of $i = 1, 2, 3, \ldots, I$ distance intervals (where $\sum_{i=1}^{I} n_i = n$); secondly, the observed number of detections or counts n_p at line or point p along with the effort data, which at bare minimum consists of the size of the search area. In case detections are made of single animals, the observed counts at the line (point) are equivalent to the number of detections at the line (point). These two types of data, distances and counts, give rise to the two components of the integrated likelihood described in this section. However, if detections are made of groups of animals (rather than single individuals), a third type of data generated from a distance sampling survey is cluster size s_e, which represents the number of individuals within the eth detected group. For simplicity, we ignore cluster sizes for this study. Methods could, however, be extended to accommodate group sizes larger than one. This may be done by considering counts of individuals (rather than detections) in the count model described below or by including a model for cluster sizes. The latter may be desirable e.g. if group size data are overdispersed (e.g. Cañadas and Hammond 2006; Schmidt et al. 2012).

In contrast to many existing covariate models for distance sampling data (e.g. Hedley and Buckland 2004; Buckland et al. 2009), the proposed integrated likelihood deals with both components of the data simultaneously—similar to e.g. Royle, Dawson, and Bates (2004) and Sillett et al. (2012). It consists of the likelihood components for the detection function, which is denoted by $L_y(\boldsymbol{\theta})$ for exact distance data (see Equation (2.5) below for interval data), and the Poisson likelihood for observed counts, $L_n(\boldsymbol{\beta}|\boldsymbol{\theta})$. We use $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ to summarise the detection function and Poisson model parameters, respectively. These are defined in more detail below. The integrated likelihood is the product of the two components:

$$L_{n,y}(\boldsymbol{\beta}, \boldsymbol{\theta}) = L_y(\boldsymbol{\theta}) L_n(\boldsymbol{\beta}|\boldsymbol{\theta}) \tag{2.1}$$

(modified from Buckland et al. 2004, Chapter 2). We consider each individual likelihood component in $L_{n,y}(\boldsymbol{\beta}, \boldsymbol{\theta})$ and begin with $L_y(\boldsymbol{\theta})$.

2.1. LIKELIHOOD COMPONENT FOR THE DETECTION FUNCTION

Let $f(y|\boldsymbol{\theta})$ denote the probability density function of observed distances which is given as

$$f(y|\boldsymbol{\theta}) = \frac{\pi(y)g(y|\boldsymbol{\theta})}{\int_0^w \pi(y)g(y|\boldsymbol{\theta})\,dy}, \tag{2.2}$$

where y is the observed distance from the line (point), and w is the truncation distance (i.e. the furthest distance from the line (point) included in the analysis, Thomas et al. 2010). $\pi(y)$ describes the expected distribution of animals with respect to the line ($\pi(y) = 1/w$) or point ($\pi(y) = 2y/w^2$). The detection function $g(y|\boldsymbol{\theta})$ may be modelled e.g. as half-normal ($g(y|\boldsymbol{\theta}) = \exp(-y^2/2\sigma^2)$, with $\boldsymbol{\theta} = \{\sigma\}$) or hazard-rate ($g(y|\boldsymbol{\theta}) = 1 - \exp(-(y/\sigma)^{-\tau})$, with $\boldsymbol{\theta} = \{\sigma, \tau\}$). The likelihood component, which is conditional on the number of detections n, may be expressed as

$$L_y(\boldsymbol{\theta}) = \prod_{e=1}^n f(y_e|\boldsymbol{\theta}), \tag{2.3}$$

where y_e refers to the eth detection (Buckland et al. 2004).

When detections are recorded in distance intervals, let f_i denote the probability that a detected animal is in the ith interval that is delineated by the cutpoints c_{i-1} and c_i:

$$f_i(\boldsymbol{\theta}) = \frac{\int_{c_{i-1}}^{c_i} f(y|\boldsymbol{\theta})\,dy}{\int_0^w f(y|\boldsymbol{\theta})\,dy}, \tag{2.4}$$

where the truncation distance, w, corresponds to the outermost cutpoint. Then, the multinomial likelihood L_{yG} replaces L_y in Equation (2.1) and may be expressed as

$$L_{yG}(\boldsymbol{\theta}) = \left(\frac{n!}{\prod_{i=1}^I n_i!}\right) \prod_{i=1}^I f_i(\boldsymbol{\theta})^{n_i}, \tag{2.5}$$

where n_i is the number of detected animals in the ith interval.

Note that in Equations (2.3) and (2.5), detections from all sites are pooled in one detection function. For modelling heterogeneity using multiple covariate distance sampling (MCDS) methods, the scale parameter σ of the half-normal or hazard-rate detection function is modelled as a function of covariates ($\sigma(z) = \delta_0 \times \exp(\sum_{q=1}^Q z_q \delta_q)$, where $\delta_q, q = 0, 1, 2, \ldots, Q$ replace σ in $\boldsymbol{\theta}$) (Marques and Buckland 2003), and the z_q represent the values of the qth covariate.

2.2. LIKELIHOOD COMPONENT FOR THE COUNT MODEL

For the log-linear Poisson model, $L_n(\boldsymbol{\beta}|\boldsymbol{\theta})$, we begin by considering a study design that consists of multiple sites each containing one or more lines (points) that are surveyed at least once. If all sites contain only one line (point) and surveys are not repeated, each count may be considered independent under conditions described by Buckland et al. (2001). However, if sites contain clusters of lines (points) and/or sites were surveyed more than once, this assumption is violated. We deal with this by grouping counts from the same site and fitting a random effect coefficient for each site in the following count model (see below in this section).

Here, we consider counts n_{jpr} at visit r to line or point p at site j as a Poisson random variable with $E(n_{jpr}) = \lambda_{jpr}$. To adjust counts for imperfect detection out to distance w, $f(y|\boldsymbol{\theta})$ from Equation (2.2) is used to estimate the effective area $\nu(\boldsymbol{\theta})$, which is defined as the area beyond which as many animals are seen as are missed within (Buckland et al. 2001). Thus, the density D_{jpr} may be expressed as

$$D_{jpr} = \lambda_{jpr}/\nu(\boldsymbol{\theta}). \tag{2.6}$$

For line transects, $\nu(\boldsymbol{\theta}) = 2l_p \int_0^w g(y|\boldsymbol{\theta})\,dy$, where l_p is the length of the line surveyed; for point transects, $\nu(\boldsymbol{\theta}) = 2\pi \int_0^w yg(y|\boldsymbol{\theta})\,dy$. These definitions for ν are given for the case where all detections are pooled in a global detection function. When modelling heterogeneity, e.g. using MCDS methods, the effective area may vary between lines (points), and the global ν becomes ν_{jpr}.

However, when replacing D_{jpr} with a covariate model ($D_{jpr} = \exp(\beta_0 + b_j + \sum_{k=1}^K x_{kjpr}\beta_k)$) and rearranging Equation (2.6), we obtain a model for the expected counts λ_{jpr}, which is now a function of the density model parameters $\boldsymbol{\beta}$ and conditional on detection function parameters $\boldsymbol{\theta}$:

$$\lambda_{jpr}(\boldsymbol{\beta}|\boldsymbol{\theta}) = \exp\left(\beta_0 + b_j + \sum_{k=1}^K x_{kjpr}\beta_k + \ln(\nu(\boldsymbol{\theta}))\right). \tag{2.7}$$

Here β_0 is the intercept, b_j the random effect for site j ($b_j \sim N(0, \sigma_b^2)$), x_k the k covariates, x_{kjpr} the covariate values measured during visit r to that line or point, and β_k the associated coefficients. The vector $\boldsymbol{\beta} = \{\beta_0, \beta_1, \beta_2, \ldots, \beta_K, \sigma_b\}$ denotes the parameters associated with the covariates affecting densities and the random effect standard deviation. Equation (2.7) is given for the general case where lines or points that may produce correlated counts, due to closeness in space and/or due to repeated measurements at the same line (point), are grouped together as site j. The inclusion of a random effect for site accommodates covariances for these measurements. However, in cases where lines (points) follow a random survey design (Buckland et al. 2001) and each line (point) is surveyed only once, the random effect term may be omitted.

Using this model for λ_{jpr}, the likelihood for the count model (Equation (2.7)), conditional on detection function parameters $\boldsymbol{\theta}$, may be expressed as

$$L_n(\boldsymbol{\beta}|\boldsymbol{\theta}) = \prod_{j=1}^{J} \int_{-\infty}^{\infty} \left(\prod_{p=1}^{P_j} \prod_{r=1}^{R_j} \frac{(\lambda_{jpr})^{n_{jpr}} \exp(-\lambda_{jpr})}{n_{jpr}!} \times \frac{1}{\sqrt{2\pi\sigma_b^2}} \exp\left(-\frac{1}{2\sigma_b^2} b_j^2\right) \right) db_j,$$

(2.8)

where J refers to the total number of sites, and P_j and R_j refer to the total number of lines (points) at and visits to the jth site, respectively. $L_n(\boldsymbol{\beta}|\boldsymbol{\theta})$ forms the second likelihood component in Equation (2.1). Note that in a maximum likelihood context the likelihood function including a random effect (for which normality is assumed) is generally formulated with an integral as shown in Equation (2.8) as the random effect is integrated out analytically (or by approximation) and the individual coefficients b_j are not estimated (e.g. McCulloch and Searle 2001). In the Bayesian context, the random effect is not integrated out analytically. Here, we use a data augmentation scheme where the individual coefficients b_j are included as parameters (or auxiliary variables) within the model and updated within the MCMC algorithm (see below).

3. THE BAYESIAN APPROACH

3.1. HIERARCHICAL MODELS

Using a Bayesian approach, random-effects models can be implemented using hierarchical models where the standard deviation of the random effect (σ_b from Equation (2.8)) is considered a random variable with a distribution rather than as a fixed value (Davison 2003). Individual random effects coefficients (b_j from Equations (2.7) and (2.8)) are fitted in the model and updated at each iteration of the chain in a Markov chain Monte Carlo (MCMC) algorithm.

Prior beliefs regarding the parameters, such as knowledge obtained from a previous study, may be included in the current study via the prior distribution. This may allow inference on model parameters in cases where too few data exist in the current study to obtain maximum likelihood estimates with great precision (e.g. Eguchi and Gerrodette 2009).

3.2. MCMC ALGORITHM

An MCMC algorithm is used to explore the posterior distribution of the parameters given the data. Commonly used MCMC methods are the Gibbs sampler and Metropolis–Hastings (MH) update. We focus on the MH update (Hastings 1970; Metropolis et al. 1953) as some of the likelihood functions that may be used to form the posterior conditional distributions of parameters are non-standard (e.g. half-normal or hazard-rate detection function that may include a covariate model for the scale parameter). In particular, we use a random walk single-update MH algorithm with normal proposal density. The proposal variance for each parameter may be obtained via pilot-tuning (Gelman, Roberts, and Gilks 1996). See Appendix A.1 for details on the MH algorithm.

3.3. Model Selection: Reversible Jump MCMC

To discriminate between competing models, we treat the model itself as a parameter and form the joint posterior distribution over both parameters and models. To explore this posterior distribution we implement an RJMCMC algorithm (Green 1995) where each iteration involves two steps; step 1: update parameters given the current model using the MH algorithm (within model move) as described above in Section 3.2; step 2: update the model using an RJ algorithm (between model move). Posterior model probabilities are estimated as the proportion of time the chain spent in a particular model after the burn-in.

For the RJ step, two main strategies may be followed. In cases where models differ only in the combination of the same set of covariates, a single RJ step may involve going through each covariate and proposing to delete or add it depending on whether it is in the current model or not. This involves generating a value for the new parameter from a proposal distribution (if we propose to add it) or setting it to zero (if we propose to delete it) and calculating the acceptance probability each time we propose to add or delete a parameter.

In those cases where all parameters of the newly proposed model change, one RJ step involves generating new values for all parameters of the new model and accepting or rejecting the new model based on the above acceptance probability. A proposed move from a half-normal detection function model to a hazard-rate model represents a simple example for this scenario. For further details of the RJMCMC algorithm, see Appendix A.2.

4. CASE STUDY

4.1. The Data

As part of a study to assess the potential benefits of herbaceous buffers around agricultural fields, Mississippi State University, Department of Wildlife, Fisheries, and Aquaculture set up a monitoring program using point transects in a number of Midwestern and Southeastern states in the US (Evans et al. 2013). Survey points located at the edge of the field were paired up: one point on a buffered treatment field and the other on a non-buffered control field of the same agricultural use and within 1–3 km of the treatment point. Each pair of points will be referred to as a site in the following. Repeat visits were made to each point during fall of three survey years (2006–2008), and each detected northern bobwhite covey was recorded along with their estimated radial distance to the point. To facilitate unbiased distance estimation, observers used satellite images of the point location and surroundings to mark each detected covey. As no estimates of covey cluster sizes were obtained, we model cluster densities (rather than densities of individuals).

Only those 11 among the 13 original states in our study were included in the analysis that contained more than 50 detections of coveys: Georgia, Iowa, Illinois, Indiana, Kentucky, Missouri, Mississippi, North Carolina, South Carolina, Tennessee and Texas. Within these states, 447 sites were visited 1–3 times in each survey year. The number of sites per state ranged from 30 to 61. After defining a truncation distance of 500 m following recommendations of Buckland et al. (2001), the analysed data included a total of 2545 detections with associated distances that were observed during 2534 counts (number of

Table 1. Lower and upper bounds for uniform prior distributions for all model parameters. The different *states* included GA, IA, IL, IN, KY, MO, MS, NC, SC, TN and TX.

Parameters	Lower	Upper
Detection Model		
Scale Intercept:	1	100000
Shape:	1	20
Year levels: 2006, 2007	−3	3
Type level: Control	−2.5	2.5
State levels: GA:TN	−2.5	2.5
Count Model		
Intercept:	−20	−7
Year levels: 2007, 2008	−1	1
Type level: Treatment	0	1
Julian Day:	−0.1	0.1
State level: IA:TX	−3	3
Random effect standard deviation	0	2

counts by state: GA 190, IA 221, IL 162, IN 217, KY 218, MO 352, MS 236, NC 244, SC 250, TN 219, TX 225).

4.2. THE BAYESIAN APPROACH

We used Equations (2.3) and (2.8) to form the integrated likelihood function as shown in (2.1). Potential covariates included in the models for $L_y(\theta)$ and $L_n(\beta|\theta)$ were the factor covariates *year* (three levels: 2006, 2007, 2008), *type* (two levels: Control or Treatment plot), *state* (11 levels) and the numeric covariate *Julian day*, which was centred around its mean before the analyses (for the $L_n(\beta|\theta)$ model only as it did not reveal any influence on detection probabilities during preliminary analyses using Distance software, Thomas et al. 2010). See Evans et al. (2013) for ecological details on modelling data from this study. Hence, eight different covariate combinations were possible for the detection function, while 16 were possible for the density model.

We assume that counts from the different points of the same sites were (positively) correlated in the same manner as repeat counts of the same point. Hence, we included one random effect coefficient per site in the model (as described in Section 2.2) where the same coefficient applied to all repeat counts at either of the two points belonging to the same site.

Uniform priors were placed on all parameters θ and β for which bounds were chosen in preliminary analyses (Table 1). Generally, this involved adding/subtracting two times the standard errors from the maximum likelihood estimate of the full model; however, bounds were extended if a parameter value reached either of these. If the parameter had natural bounds, e.g. zero for the lower bound of the random effect standard deviation, these were adopted.

To make summary statistics of parameters directly comparable to the maximum likelihood approach (see Section 4.3), the last covariate levels (in numerical or alphabetic order) of detection function parameters were absorbed in the intercept to follow the parameterisation of factor levels in Distance software. Likewise, the first levels were absorbed in

the intercept for the count model to follow parameterisation of factor levels in the *glmer* function from the *lme4* package in R.

Preliminary investigation of the distance data indicated that the hazard-rate detection function provided a much better fit than the half-normal. Hence, we included eight different hazard-rate models as choices for the probability density function of observed distances $f(y|\theta)$ in $L_y(\theta)$: one global (with no covariates) and seven multiple covariate models. For the global model, only the scale and the shape parameters required estimation (see Section 2 for details). The multiple covariate models contained additional parameters as the scale parameter was modelled as a function of one, two or three of the covariates.

For $L_n(\boldsymbol{\beta}|\boldsymbol{\theta})$, λ_{jpr} from Equation (2.7) was modelled including a intercept and combinations of the four covariates as well as a random effect for site. In preliminary analyses, we investigated whether the count data were overdispersed. We fitted a *glmm* with a quasipoisson distribution using the full models for both detection and counts to calculate the offset. A quasi-Poisson *glmm* can be fitted using the *gamm* function of the *mgcv* package. The estimate of the dispersion parameter was 1.10. Hence, we assumed that Poisson was appropriate to use.

The chain was started without any covariates for the detection and count models. During a single RJ step of each iteration, each of the covariates was proposed to be added or deleted depending on whether it was in the current model or not. The values u for the new parameters contained in the new model were drawn from parameter-specific proposal distributions shown in Table 2. These were initially defined as normal distributions with mean and standard deviation equal to the maximum likelihood estimates and standard errors from the full models; however, we adjusted means (by averaging estimates across different models for the respective parameters) and standard deviations during pilot-tuning to improve model mixing.

To move from e.g. a global hazard-rate model to a model including a covariate, the global scale parameter σ was converted into $\delta_0 \times \exp(z_1\delta_1)$ with $\sigma = \delta_0$ and $u = \delta_1$, where δ_1 is the coefficient associated with covariate z_1. The bijective function in this case (as well as in all the other possible model moves) was the identity function similar to the example shown in Appendix A.2. Therefore, the Jacobian $|J|$ (from Equation (A.4) in Appendix A.2) equalled one. We assume that all models were equally likely a priori, hence the probability of moving to model m conditional on the chain being in model m', $P(m|m')$, was equal to $P(m'|m)$ and vice versa for all possible model moves and cancelled when calculating the acceptance probability (see Equation (A.4)).

Proposal distributions for the MH step were normal where the mean was the current value of the parameter and the standard deviation was parameter-specific. The RJ and MH step together completed one iteration. A total of 100,000 iterations were carried out where the first 10,000 were considered the burn-in period and ignored when obtaining model probabilities and summary statistics for parameters. Visual inspection of raw trace plots from different starting points for parameters suggested that convergence had been achieved within 10,000 iterations.

We conducted a prior sensitivity analysis where the uniform priors were replaced with normal priors for all parameters. We also tested the robustness of our estimated model

Table 2. Mean and standard deviation (SD) of Normal proposal distributions for parameters proposed to be
added or deleted during the RJ step of the RJMCMC algorithm. All parameters were categorical, except
for continuous Julian day.

Parameters	Mean	SD
Detection Model		
Year level 2006	0.11	0.10
Year level 2007	−0.15	0.10
Type level: Control	0.50	0.10
State level: GA	0.42	0.10
State level: IA	0.21	0.10
State level: IL	0.70	0.10
State level: IN	0.67	0.10
State level: KY	0.64	0.10
State level: MO	0.69	0.10
State level: MS	0.61	0.10
State level: NC	0.66	0.10
State level: SC	0.03	0.10
State level: TN	0.47	0.10
Count Model		
Year level: 2007	0.16	0.05
Year level: 2008	0.08	0.05
Type level: Treatment	0.42	0.10
Julian Day:	−0.01	0.01
State level: IA	0.71	0.24
State level: IL	−0.49	0.24
State level: IN	−1.16	0.23
State level: KY	−0.41	0.22
State level: MO	0.01	0.20
State level: MS	−0.38	0.22
State level: NC	−1.36	0.23
State level: SC	0.07	0.22
State level: TN	−1.05	0.23
State level: TX	1.77	0.21

probabilities by starting the RJMCMC algorithm from the full detection and count models
(as opposed to the intercept only models). Both analyses revealed nearly identical results
as those presented in Section 4.4. Hence, we were confident that our results had converged.

4.3. THE CLASSICAL APPROACH

To compare the Bayesian approach with the classical, the data were analysed using
the two-stage approach (Buckland et al. 2009), extended to include a random effect for
site in the count model (Oedekoven et al. 2013). The first step included fitting a detection
function to observed distances by maximising the likelihood in Equation (2.3). The same
eight hazard-rate models were explored as in the Bayesian approach, i.e. global and MCDS
models with combinations of the covariates *state*, *type* and *year*. The effective area v_{jpr}
was estimated using the best model for $f(y|\boldsymbol{\theta})$.

In a second step, the effective area was incorporated into the count model for λ_{jpr} and
parameter estimates obtained using the *glmer* function of the *lme4* package (Bates 2009b)

Table 3. Models and their probabilities resulting from RJMCMC and bootstrap analyses. Each density model included an intercept and a random effect for site in addition to shown covariates (JD = Julian day). Model probabilities refer to the percentage the respective models were chosen during 90 000 iterations (after 10 000 iterations of burn-in) for RJMCMC and during 999 bootstrap iterations.

Model	RJMCMC	Two-Stage
Detection Model		
MCDS: Type	<0.001	–
MCDS: State	–	0.01
MCDS: Year + State	–	0.16
MCDS: Type + State	<0.001	0.02
MCDS: Year + Type + State	1.00	0.81
Count Model		
Type + State	–	<0.005
Year + Type + State	–	0.01
Type + JD + State	0.89	0.10
Year + Type + JD + State	0.11	0.89

in R. The likelihood for the *glmer* function is equivalent to Equation (2.8), except that the random effect is integrated out using an approximation for the integral (Bates 2009a). The same 16 models were explored as in the Bayesian approach, including a fixed intercept and a random effect for site and combinations of the four covariates *state*, *type*, *year* and *Julian day*.

Best fitting models for both steps were found by minimum AIC values. As the effective area represents an estimate but is included in the model as if it was a known constant, non-parametric bootstrapping was used to estimate uncertainty (bootstrap standard errors (BSE) and 95 % confidence intervals) of parameter estimates. To implement a non-parametric bootstrap routine with 999 repeats, an automatic model selection was set up in R that included calls to the MCDS engine from the Distance software (Thomas et al. 2010) for the first step. For each bootstrap iteration, sites were resampled with replacement until the original number of sites was obtained (Buckland et al. 2009). To include model uncertainty in inference, the strategy followed was to select best fitting models based on minimum AIC values for each bootstrap iteration (Buckland, Burnham, and Augustin 1997).

4.4. RESULTS

In the following we refer to those models with highest probabilities as the preferred models. For the Bayesian approach, the preferred detection function model included the covariates *year*, *type* and *state* in the model for the scale parameter of the hazard-rate key function (probability = 1.00 to two decimal places, Table 3). Two other models were visited within the RJMCMC algorithm with probabilities of <0.001 that included two (*type* and *state*) or one covariate only (*type*). The same model with all three covariates was the preferred model for the two-stage approach having been selected by AIC in 81 % of bootstrap resamples. Three other models were selected: one with covariates *year* and *state* (16 %), one with *type* and *state* (2 %) and one with *state* alone (1 %).

For the count model, two models dominated the RJMCMC algorithm, the model with covariates *type*, *Julian day* and *state* as the preferred model (0.89 probability) and the full

model (*year + type + Julian day + state*, 0.11 probability, Table 3). For the bootstrap, the latter was the preferred model, selected in 89 % of resamples, while the former was the second most frequently chosen model (10 %). Two other models were chosen during the bootstrap, the model including the covariates *year*, *type* and *state* (1 %) and the model including covariates *type* and *state* (<1 %). Hence, the largest discrepancy in model probabilities between the two analysis methods was with regard to covariate *year* for which the total probabilities to be included in any model was 0.11 for the RJMCMC algorithm and 0.90 for the bootstrap (Table 3). However, 95 % confidence intervals obtained from the bootstrap overlapped zero for both *year* coefficients (Table 4), indicating that this covariate might have less importance than suggested by model probabilities for the bootstrap.

For the parameters of the detection function model, the posterior means of the parameters in the preferred model were in most cases similar to the maximum likelihood estimates resulting from the two-stage analysis of the original data (Table 4). The intercept for the scale parameter and the shape parameter were larger for the Bayesian approach, while the coefficients for the scale parameter were on average smaller. Histograms of detections by state and estimated expected probability density functions (pdf) of observed distances from the Bayesian approach are shown in Fig. 1. States SC and TX had the lowest average detection probabilities as indicated by the smallest coefficients for these states (Table 4). For these states, the steep decline after peak values for the pdf indicated rapidly declining detection probabilities beyond ∼100 m (Fig. 1).

Interestingly, measures of uncertainty were mostly smaller for the Bayesian approach despite the fact that both stages from the two-stage approach were combined in one. The posterior standard deviations were smaller than the bootstrap standard errors for all detection function parameters. The 95 % credible intervals were narrower than the 95 % confidence intervals for all but four detection function parameters (state coefficients IN, MS, NC and TN). Intervals from the two approaches overlapped in all cases for the detection function parameters.

For the count model, means and intervals were again similar between the two approaches (Table 4). However, slight discrepancies in means for count model coefficients existed, which might have been due to that the best model from the two-stage approach contained the additional covariate *year* and/or to Monte Carlo error. Further reasons are discussed in Section 5. Also, measures of uncertainty were again mostly smaller for the Bayesian approach: standard deviations from the Bayesian approach were smaller for all covariates in the count model compared to BSEs. The 95 % credible intervals were narrower for all coefficients of the count model compared to the 95 % confidence intervals, except for the covariate *Julian day* where they were equal. The 95 % credible and confidence intervals overlapped for all count model parameters. The only covariate selected for the preferred count model for the two-stage approach that was not also in the preferred model for the Bayesian approach was *year*. The 95 % confidence intervals for both *year* coefficients included zero, indicating that this covariate might have been negligible for the count model.

The parameter of interest in these models was the coefficient for the level Treatment of the *type* covariate in the count model. This was 0.62 (SD = 0.07) and 0.63 (BSE =

Table 4. Mean, standard deviation (SD) and 95 % credible intervals (CRI) from the RJMCMC analysis along with maximum likelihood estimates (MLE), bootstrap standard errors (BSE) and 95 % confidence intervals (CI) using the two-stage approach for the models with the highest probabilities (see Table 3 for model probabilities). Units of measurements were metres for the detection function model and square metres for the count model.

	RJMCMC			Two-Stage		
	Mean	SD	95 % CRI	MLE	BSE	95 % CI
Detection Model						
Scale Intercept	152.96	8.97	135.47, 170.12	138.59	16.13	112.26, 163.79
Shape	3.30	0.16	3.00, 3.63	3.01	0.27	2.68, 3.41
Scale: Year 2006	0.06	0.03	0.01, 0.12	0.10	0.06	−0.05, 0.14
Scale: Year 2007	−0.11	0.04	−0.19, −0.05	−0.15	0.05	−0.25, -0.1
Scale: Type Control	0.13	0.04	0.05, 0.20	0.15	0.05	0.05, 0.23
Scale: State GA	0.38	0.07	0.23, 0.52	0.42	0.17	0.05, 0.54
Scale: State IA	0.17	0.10	−0.01, 0.36	0.21	0.16	−0.12, 0.30
Scale: State IL	0.66	0.09	0.48, 0.85	0.70	0.17	0.35, 0.76
Scale: State IN	0.62	0.10	0.43, 0.81	0.66	0.14	0.34, 0.72
Scale: State KY	0.58	0.08	0.43, 0.74	0.64	0.12	0.35, 0.68
Scale: State MO	0.62	0.06	0.51, 0.73	0.69	0.09	0.46, 0.71
Scale: State MS	0.55	0.07	0.41, 0.70	0.61	0.10	0.37, 0.64
Scale: State NC	0.60	0.09	0.43, 0.79	0.66	0.12	0.35, 0.70
Scale: State SC	0.01	0.08	−0.14, 0.16	3E−5	0.14	−0.29, 0.12
Scale: State TN	0.44	0.10	0.25, 0.63	0.47	0.12	0.19, 0.54
Count Model: Random Effects						
Standard deviation	0.82	0.05	0.73, 0.91	0.78	0.04	0.69, 0.81
Count Model: Coefficients						
Intercept Density	−13.10	0.18	−13.43, −12.73	−13.23	0.33	−13.91, −12.87
Year 2007	–	–	–	0.17	0.13	−0.16, 0.37
Year 2008	–	–	–	0.17	0.11	−0.12, 0.31
Type Treatment	0.62	0.07	0.48, 0.75	0.63	0.12	0.36, 0.71
Julian Day	−0.01	2E−3	−0.02, −0.01	−0.01	3E−3	−0.02, −0.01
State IA	−0.81	0.29	−1.38, −0.23	−0.74	0.44	−1.65, −0.24
State IL	−0.59	0.27	−1.12, −0.06	−0.53	0.38	−1.25, −0.07
State IN	−1.24	0.27	−1.79, −0.71	−1.18	0.41	−1.99, −0.70
State KY	−0.47	0.25	−0.98, 0.03	−0.44	0.34	−1.07, −0.02
State MO	0.01	0.22	−0.44, 0.42	0.05	0.34	−0.63, 0.46
State MS	−0.43	0.25	−0.92, 0.05	−0.37	0.34	−1.04, 0.05
State NC	−1.39	0.26	−1.88, −0.87	−1.31	0.36	−1.99, −0.87
State SC	0.01	0.27	−0.53, 0.53	0.08	0.42	−0.76, 0.56
State TN	−1.10	0.28	−1.65, −0.57	−1.03	0.38	−1.80, −0.60
State TX	1.74	0.18	1.33, 1.99	1.46	0.29	0.99, 1.81

0.12) for the Bayesian and the two-stage approach, respectively, indicating an increase in covey densities on treatment plots by 85 % ($E[\exp(type\ coefficient)] = 1.85$) or 88 % ($\exp(0.63) = 1.88$) by the respective methods.

As described above, the expected densities can be calculated using the values of the β parameters from the count model. These values are obtained from the posterior distribution of these parameters for the Bayesian approach or maximum likelihood estimates for the two-stage approach (see above Equations (2.6) and (2.7) for details). For calculating baseline estimates of the expected covey densities for the RJMCMC algorithm and two-stage approach, we used those iterations from the respective methods where the preferred
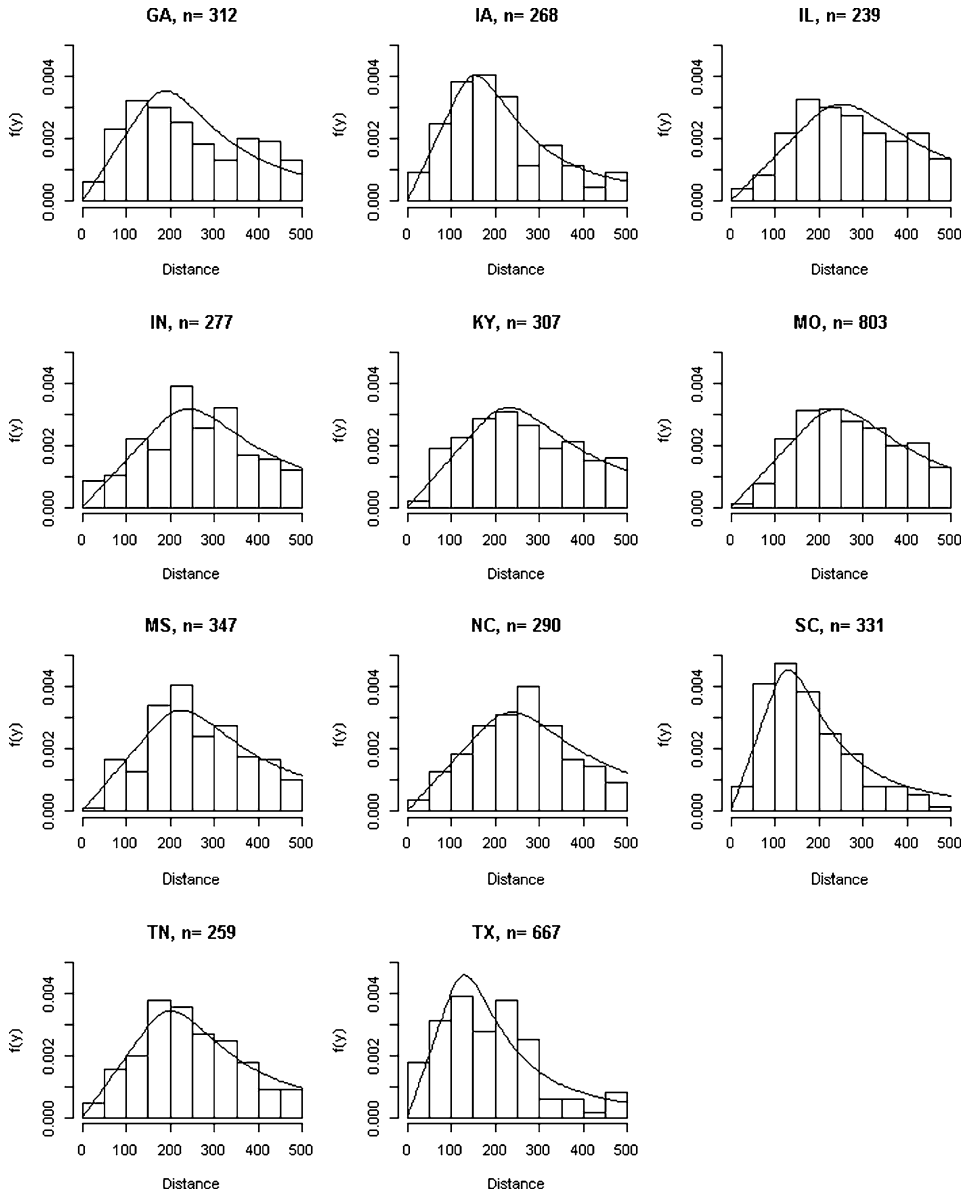
Figure 1. Estimated probability density functions, $f(y)$ (y-axis), using mean parameter values from the Bayesian approach (Table 4) against radial distance from the point (x-axis, in metres), and histograms of detections for each of 11 states (GA, IA, IL, IN, KY, MO, MS, NC, SC, TN, TX); here, n represents the number of detections for the respective states. For each state, $f(y)$ was averaged over all levels of covariates *year* and *type*.

count model was chosen, excluding the burn-in iterations for the RJMCMC. Using Equation (2.7), we set the covariates to those levels that were absorbed by the intercept of the density model, i.e. *year* = 2006, *type* = Control, *Julian day* = 0 (which is equivalent to its mean as we centred the data for this covariate) and *state* = GA and added a random effect contribution ($0.5 \times \sigma_b^2$ replaces b_j from Equation (2.7)) when calculating the aver-
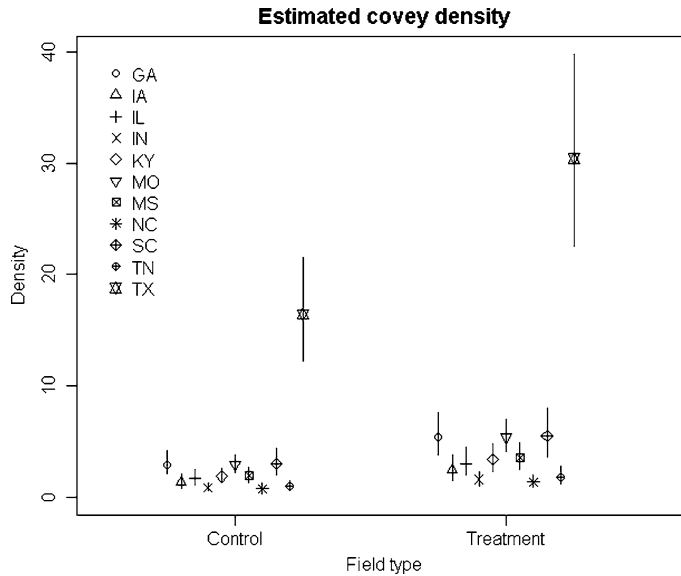
Figure 2.   Estimated covey densities (coveys per km^2) and 95 % CRIs for each of 11 states shown for unbuffered control and buffered treatment fields.

age expected density across all sites). The estimated baseline expected density from the RJMCMC algorithm was 2.91 coveys per km^2 (SD = 0.55, 95 % CRI = (2.05, 4.14)). The posterior distribution of the expected density was right-skewed with a mode of 2.60 coveys per km^2. The estimated expected density resulting from the two-stage approach was 2.44 coveys per km^2 (BSE = 0.88, 95 % CI = (1.25, 4.66)). The estimated expected densities for each *state* and *type* combination are shown in Fig. 2.

5. DISCUSSION

There are two main aspects in this paper that are innovative and deserve comparison to existing methods. We present a novel approach for combining the likelihood functions for analysing distance sampling data in Section 2, which is easily applicable to both interval and exact distance data. We also present a Bayesian approach for analysing distance sampling data of multiple types in a straightforward manner. Different detection functions (the half-normal, hazard-rate or others) may easily be implemented. It may also be extended to include adjustment terms (added to the half-normal or hazard-rate model, Buckland et al. 2001) or covariates in the shape parameter. We provide the R code as online supplementary material, which is annotated for easy adaptation.

Bayesian methods have been used before for analysing line transect data with a global half-normal detection function (e.g. Royle and Dorazio 2008, Chapter 7.1; Eguchi and Gerrodette 2009; Gimenez et al. 2009; Zhang 2011), a half-normal with covariates (e.g. Gerrodette and Eguchi 2011; Moore and Barlow 2011) or a hazard-rate with one covariate (Schmidt et al. 2012). Our approach allows simultaneous exploration of model and parameter space including different detection functions and different covariate combinations

for both the detection and count models via an RJMCMC algorithm. Conn, Laake, and Johnson (2012) described an RJMCMC algorithm for distance sampling data. However, in their case, the RJ step refers to adding/deleting unobserved animals as part of the data augmentation and not to exploring different models for the detection function or counts.

The log-linear Poisson model for counts described in Section 2 does not depend on a random survey design in contrast with the classical distance sampling approach and data arising from surveys conducted from platforms of opportunity may be used (Hedley and Buckland 2004). It allows identification of relationships between abundance or density and parameters of interest, such as the *type* covariate in our case study, which may be of interest for designed experiments and wildlife management studies (see also Gerrodette and Eguchi 2011). Similar to Hedley and Buckland (2004) and Buckland et al. (2009), our count model may be extended to include smooth functions for continuous covariates, e.g. by fitting polynomial splines using the B-spline basis, or the Poisson likelihood may be replaced with a negative binomial likelihood if more appropriate, e.g. in case overdispersion in the count data is present (e.g. Royle, Dawson, and Bates 2004, Sillett et al. 2012).

Our approach is similar to Hedley and Buckland (2004) in that we use the conditional formulation of the probability density function of observed distances (Buckland et al. 2001) and information on detection probabilities enters the count model via an offset; we model counts as *density × effective area*. However, Hedley and Buckland (2004) or Buckland et al. (2009) analyse their data in two stages. In their second stage count model, they condition on the estimate of the effective area derived from the first stage detection model. This requires conducting non-parametric bootstrapping so that uncertainty associated with estimating the detection function (and the effective area) propagates into the second stage count model. Our integrated likelihood approach estimates all parameters simultaneously, allowing direct quantification of the precision of the parameters in the count model with taking proper account of the estimation of detection function parameters.

Using the Poisson model with a random effect for estimating densities as defined in Equation (2.7) also allows us to accommodate correlated measurements due to closeness in space and/or time, as occurs e.g. when there are repeat counts at the same line (point). This differs from the integrated likelihood described by Royle, Dawson, and Bates (2004). These authors considered the true but unknown abundances at the site as a random effect with a Poisson distribution (in their notation $N_i \sim Poisson(\lambda_i)$) and integrate it out by summation. They derive a Poisson likelihood for the observed counts with expected value equal to $\lambda_i \pi_k(\theta)$, where $\pi_k(\theta)$ is obtained using the unconditional probability density function of observed distances and describes the probability that an animal occurs and is detected in the kth distance interval. Hence, these authors model counts as *abundance N × detected proportion of N* for each interval. Note that we use the term *integrated likelihood* in the spirit of integrating the likelihood components pertaining to two different data types (described in Section 2), while Royle, Dawson, and Bates (2004) use the same term in the context of integrating out a nuisance parameter (although they combine the likelihood components from two different data sources as well).

In contrast to Royle, Dawson, and Bates (2004), we model variations in observed counts between the different sites (those variations not explained by any of the fixed effects in-

cluded in the model) as a normally distributed random effect with mean zero, hence accounting for correlations between measurements at the same sites. This represents an extension to Royle, Dawson, and Bates who include one count per site in the analysis. With the inclusion of site random effects, our approach allows us to incorporate repeat counts from the same sites in the analysis. Our approach assumes that all counts from the same site are positively correlated and only requires estimation of one additional parameter, the random effect standard deviation. Chelgren et al. (2011), on the other hand, extended the approach of Royle, Dawson, and Bates (2004) by including a random effect for plot by week in the abundance model which requires estimation of week-specific variance parameters. However, including random effects allows inference on the wider area that these sites represent and to obtain unbiased estimates of coefficients retained in the count model. Bias in coefficient estimates may occur for example if some sites with high bird densities were visited more frequently than those with low densities, and this variation was not modelled as a fixed or random effect.

The comparison of summary statistics for model parameters from the Bayesian approach with those from the two-stage approach revealed some differences in means and point estimates (Table 4) which cannot be due to prior sensitivity as we used uniform priors on all parameters for the Bayesian approach. We assume that these differences may have been due to the fact that—as opposed to the two-stage approach—the likelihoods for both components of our model were combined for the integrated likelihood and influence each other. We argue, in concurrence with Johnson, Laake, and Ver Hoef (2010), that simultaneous estimation of all parameters in one stage represents a more realistic model without having to rely on the assumption of a true detection function model. Whether the smaller uncertainty estimates from the Bayesian approach compared to the two-stage approach were specific to our case study or can be expected in general is beyond the scope of this paper.

Our Bayesian approach provides improvements over previous approaches. Besides the often stated benefits for Bayesian analyses, e.g. allowing for prior information to be included, our Bayesian approach provided a particular benefit for using the integrated likelihood defined in Section 2: it might be challenging in some cases, such as our case study, to find the maximum likelihood estimates for all parameters in one step. The covey data included a total of 2545 observed distances during 2534 counts, and the full model included 31 parameters with a random effect (447 sites). Using maximum likelihood methods, the random effect coefficients are not estimated individually but are integrated out during the optimisation of the likelihood. However, due to the integrated nature of the detection and density models, functions such as *glmer* from the *lme4* package in R may not be used as these treat the offset as a constant. Using the hierarchical model set up for the Bayesian approach, the random effect coefficients are included in the model specification and updated during each iteration. Due to this data augmentation method, no numerical integration is necessary providing a straightforward technique to explore the parameter space.

Bayesian methods also offer efficient exploration of model space with the use of RJM-CMC. By contrast, using a maximum likelihood approach, a model selection routine that considered all possible model combinations for our case study would have required maximising $L_{y,n}(\boldsymbol{\beta}, \boldsymbol{\theta})$ for 128 models (possible combinations of eight detection functions and

16 count models). RJMCMC, on the other hand, allows incorporating model uncertainty into a single chain.

APPENDIX

A.1 Metropolis–Hastings

We use a single-update random walk MH algorithm where we cycle through each parameter in $L_{n,y}(\boldsymbol{\beta}, \boldsymbol{\theta})$. To use a simple scenario, assume that $\boldsymbol{\beta} = \{\beta_0, \sigma_b\}$. Then, e.g. for parameter β_0 with current value β_0^t, we propose to move to a new state, β_0', with $\beta_0' \sim (\beta_0^t, \sigma_{\beta_0}^2)$. This newly proposed state is accepted as the new state with probability $\alpha(\beta_0'|\beta_0^t)$ given by

$$\alpha\big(\beta_0' \mid \beta_0^t\big) = \min\left(1, \frac{L_{n,y}(\beta_0', \sigma_b^t, \boldsymbol{\theta}^t) p(\beta_0') q(\beta_0^t|\beta_0')}{L_{n,y}(\beta_0^t, \sigma_b^t, \boldsymbol{\theta}^t) p(\beta_0^t) q(\beta_0'|\beta_0^t)}\right). \tag{A.1}$$

Here, $q(\beta_0'|\beta_0^t)$ denotes the proposal density of β_0' given the current state is β_0^t. We note that the terms $q(\beta_0^t|\beta_0')$ and $q(\beta_0'|\beta_0^t)$ cancel in the acceptance probability since we use a symmetrical proposal distribution. The analogous MH updates are used for random effect coefficients. Proposal variances are chosen via pilot-tuning.

A.2 Model Selection: Reversible Jump MCMC

The joint posterior distribution of models and parameters is given (up to proportionality) by

$$\pi_{n,y}(\boldsymbol{\beta}_m, \boldsymbol{\theta}_m, m) \propto L_{n,y}(\boldsymbol{\beta}_m, \boldsymbol{\theta}_m, m) p(\boldsymbol{\beta}_m, \boldsymbol{\theta}_m|m) p(m), \tag{A.2}$$

where $L_{n,y}(\boldsymbol{\beta}_m, \boldsymbol{\theta}_m, m)$ denotes the probability density function of the data given current parameter values $\boldsymbol{\beta}_m$ and $\boldsymbol{\theta}_m$ and model m, $p(\boldsymbol{\beta}_m, \boldsymbol{\theta}_m|m)$ the prior distribution for model parameters $\boldsymbol{\beta}_m$ and $\boldsymbol{\theta}_m$, and $p(m)$ the prior probability of model m. The RJMCMC algorithm is used to explore the parameter and model space simultaneously (Green 1995).

Each iteration involves two steps: a within model move and a between model move. During the within model move, the Metropolis–Hastings (MH) algorithm is used to update the parameters given the model (as described above in Section 3.2). During the between model move, the reversible jump (RJ) step, model m conditional on the current parameter values is updated. This move involves a proposal to update the model itself; suppose the chain is in model m and we propose to move to model m'. A bijective function describes the relationship between the current and proposed parameters and is used to convert parameters from model m to parameters for model m'. In a simple scenario, say, where model m contains parameters $\boldsymbol{\beta} = \{\beta_0, \beta_1\}$ and model m' contains parameters $\boldsymbol{\beta}' = \{\beta_0', \beta_2'\}$, the bijective function might be expressed as an identity function:

$$\beta_0' = \beta_0 \qquad u' = \beta_1 \qquad \beta_2' = u. \tag{A.3}$$

Here u and u' are random samples from some proposal distributions for the respective parameters. The acceptance probability may then be expressed as

$$A = \frac{\pi_{n,y}(\boldsymbol{\beta}', m') P(m|m') q'(u')}{\pi_{n,y}(\boldsymbol{\beta}, m) P(m'|m) q(u)} |J|, \tag{A.4}$$

where $P(m'|m)$ denotes the probability of proposing to move to model m' given that the chain is in model m, $q(u)$ and $q'(u')$ are the proposal densities of u and u', and $|J|$ is the Jacobian (which equals one if the bijective function is the identity function).

ACKNOWLEDGEMENTS

REFERENCES

Bates, D. (2009a), "Adaptive Gauss–Hermite Quadrature for Generalized Linear or Nonlinear Mixed Models. R Package Version 0.999375-31," Technical Report. http://lme4.r-forge.r-project.org/.

Bates, D. (2009b), "Computational Methods for Mixed Models. R Package Version 0.999375-31," Technical Report. http://lme4.r-forge.r-project.org/.

Buckland, S. T., Burnham, K. P., and Augustin, N. H. (1997), "Model Selection: An Integral Part of Inference," *Biometrics*, 53 (2), 603–618.

Buckland, S. T., Goudie, I. B. J., and Borchers, D. L. (2000), "Wildlife Population Assessment: Past Developments and Future Directions," *Biometrics*, 56, 1–12.

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., and Thomas, L. (2001), *Introduction to Distance Sampling*, London: Oxford University Press.

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., and Thomas, L. (2004), *Advanced Distance Sampling*, London: Oxford University Press.

Buckland, S. T., Russell, R. E., Dickson, B. G., Saab, V. A., Gorman, D. G., and Block, W. M. (2009), "Analysing Designed Experiments in Distance Sampling," *Journal of Agricultural, Biological, and Environmental Statistics*, 14, 432–442.

Cañadas, A., and Hammond, P. S. (2006), "Model-Based Abundance Estimates for Bottlenose Dolphins off Southern Spain: Implications for Conservation and Management," *Journal of Cetacean Research and Management*, 8 (1), 13–27.

Chelgren, N. D., Samora, B., Adams, M. J., and McCreary, B. (2011), "Using Spatiotemporal Models and Distance Sampling to Map the Space Use and Abundance of Newly Metamorphosed Western Toads (Anaxyrus Boreas)," *Herpetological Conservation and Biology*, 6 (2), 175–190.

Conn, P. B., Laake, J. L., and Johnson, D. S. (2012), "A Hierarchical Modeling Framework for Multiple Observer Transect Surveys," *PLoS ONE*, 7 (8), e42294.

Davison, A. C. (2003), *Statistical Models*, Cambridge: Cambridge University Press.

Durban, J. W., and Elston, D. A. (2005), "Mark-Recapture with Occasion and Individual Effects: Abundance Estimation Through Bayesian Model Selection in a Fixed Dimensional Parameter Space," *Journal of Agricultural, Biological, and Environmental Statistics*, 10 (3), 291–305.

Eguchi, T., and Gerrodette, T. (2009), "A Bayesian Approach to Line-Transect Analysis for Estimating Abundance," *Ecological Modelling*, 220, 1620–1630.

Evans, K. O., Burger, L. W., Oedekoven, C. S., Smith, M. D., Riffell, S. K., Martin, J. A., and Buckland, S. T. (2013), "Multi-Region Response to Conservation Buffers Targeted for Northern Bobwhite," *The Journal of Wildlife Management*, 77, 716–725.

Gelman, A., Roberts, G. O., and Gilks, W. R. (1996), "Efficient Metropolis Jumping Rules," in *Bayesian Statistics*, Vol. 5, eds. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford: Oxford University Press, pp. 599–608.

Gerrodette, T., and Eguchi, T. (2011), "Precautionary Design of a Marine Protected Area Based on a Habitat Model," *Endangered Species Research*, 15 (2), 159–166.

Gimenez, O., Bonner, S. J., King, R., Parker, R. A., Brooks, S. P., Jamieson, L. E., Grosbois, V., Morgan, B. J. T., and Thomas, L. (2009), "WinBUGS for Population Ecologists: Bayesian Modeling Using Markov Chain Monte Carlo Methods," in *Modeling Demographic Processes in Marked Populations. Environmental and Ecological Statistics*, Vol. 3, eds. D. L. Thomson, E. G. Cooch, and M. J. Conroy, Berlin: Springer, pp. 883–915.

Green, P. J. (1995), "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination," *Biometrika*, 82 (4), 711–732.

Hastings, W. K. (1970), "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, 57 (1), 97–109.

Hedley, S. L., and Buckland, S. T. (2004), "Spatial Models for Line Transect Sampling," *Journal of Agricultural, Biological, and Environmental Statistics*, 9, 181–199.

Johnson, D. S., Laake, J. L., and Ver Hoef, J. M. (2010), "A Model-Based Approach for Making Ecological Inference from Distance Sampling Data," *Biometrics*, 66, 310–318.

Karunamuni, R. J., and Quinn II., T. J. (1995), "Bayesian Estimation of Animal Abundance for Line Transect Sampling," *Biometrics*, 51, 1325–1337.

King, R., Morgan, B. J. T., Gimenez, O., and Brooks, S. P. (2010), *Bayesian Analysis for Population Ecology*, London/Boca Raton: Chapman & Hall/CRC Press.

Marcot, B. G., Holthausen, R. S., Raphael, M. G., Rowland, M. M., and Wisdom, M. J. (2001), "Using Bayesian Belief Networks to Evaluate Fish and Wildlife Population Viability Under Land Management Alternatives from an Environmental Impact Statement," *Forest Ecology and Management*, 153, 29–42.

Marques, F. F. C., and Buckland, S. T. (2003), "Incorporating Covariates Into Standard Line Transect Analyses," *Biometrics*, 53, 924–935.

McCulloch, C. E., and Searle, S. R. (2001), *Generalized, Linear, and Mixed Models*, New York: Wiley.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953), "Equations of State Calculations by Fast Computing Machines," *Journal of Chemical Physics*, 21 (6), 1087–1091.

Moore, J. E., and Barlow, J. (2011), "Bayesian State-Space Model of Fin Whale Abundance Trends from a 1991–2008 Time Series of Line-Transect Surveys in the California Current," *Journal of Applied Ecology*, 48 (5), 1195–1205.

Oedekoven, C. S., Buckland, S. T., Mackenzie, M. L., Evans, K. O., and Burger, L. W. (2013), "Improving Distance Sampling: Accounting for Covariates and Non-independency Between Sampled Sites," *Journal of Applied Ecology*, 50 (3), 786–793.

Royle, J. A., and Dorazio, R. M. (2008), *Hierarchical Modeling and Inference in Ecology: The Analysis of Data from Populations, Metapopulations and Communities*, San Diego: Academic Press.

Royle, J. A., Dawson, D. K., and Bates, S. (2004), "Modelling Abundance Effects in Distance Sampling," *Ecology*, 85 (6), 1591–1597.

Schmidt, J. H., Lindberg, M. S., Johnson, D. S., Conant, B., and King, J. (2009), "Evidence of Alaskan Trumpeter Swan Population Growth Using Bayesian Hierarchical Models," *The Journal of Wildlife Management*, 73 (5), 720–727.

Schmidt, J. H., Rattenbury, K. L., Lawler, J. P., and MacCluskie, M. C. (2012), "Using Distance Sampling and Hierarchical Models to Improve Estimates of Dall's Sheep Abundance," *The Journal of Wildlife Management*, 76 (2), 317–327.

Sillett, T. S., Chandler, R. B., Royle, J. A., Kéry, M., and Morrison, S. A. (2012), "Hierarchical Distance-Sampling Models to Estimate Population Size and Habitat-Specific Abundance of an Island Endemic," *Ecological Applications*, 22, 1997–2006.

Thomas, L., Buckland, S. T., Rexstad, E. A., Laake, J. L., Strindberg, S., Hedley, S. L., Bishop, J. R. B., Marques, T. A., and Burnham, K. P. (2010), "Distance Software: Design and Analysis of Distance Sampling Surveys for Estimating Population Size," *Journal of Applied Ecology*, 47, 5–14.

Zhang, S. (2011), "On Parametric Estimation of Population Abundance for Line Transect Sampling," *Environmental and Ecological Statistics*, 18, 79–92.