Tigers on Trails: Occupancy Modeling for Cluster Sampling

Hines, J.E.[1,4], J.D. Nichols[1], J.A. Royle[1], D.I. MacKenzie[2], A. M. Gopalaswamy[3], N. Samba Kumar[3] and K.U. Karanth[3]

[1] *United States Geological Survey, Patuxent Wildlife Research Center, Laurel Maryland, 20708, USA.*

[2] *Proteus Wildlife Research Consultants, P.O. Box 5193, Dunedin, New Zealand*

[3] *Wildlife Conservation Society – India Program, Centre for Wildlife Studies, Bangalore, Karnataka-560042, India*

[4] *Email: jhines@usgs.gov*

Abstract: Occupancy modeling focuses on inference about the distribution of organisms over space, using temporal or spatial replication to permit inference about the detection process. Inference based on spatial replication strictly requires that replicates be selected randomly and with replacement, but the importance of these design requirements is not well understood. This paper focuses on an increasingly popular sampling design based on spatial replicates that are not selected randomly and that are expected to exhibit Markovian dependence. We develop two new occupancy models for data collected under this sort of design, one based on an underlying Markov model for spatial dependence and the other based on a trap response model with Markovian detections. We then simulated data under the model for Markovian spatial dependence and fit the data to standard occupancy models and to the two new models. Bias of occupancy estimates was substantial for the standard models, smaller for the new trap response model and negligible for the new spatial process model. We also fit these models to data from a large scale tiger occupancy survey recently conducted in Karnataka State, southwestern India. In addition to providing evidence of a positive relationship between tiger occupancy and habitat, model selection statistics and estimates strongly supported the use of the model with Markovian spatial dependence. This new model provides another tool for the decomposition of the detection process, which is sometimes needed for proper estimation and which may also permit interesting biological inferences. In addition to designs employing spatial replication, we note the likely existence of temporal Markovian dependence in many designs using temporal replication. The models developed here will be useful either directly, or with minor extensions, for these designs as well. We believe that these new models represent important additions to the suite of modeling tools now available for occupancy estimation in conservation monitoring. More generally, this work represents a contribution to the topic of cluster sampling for situations in which there is a need for specific modeling (e.g., reflecting dependence) for the distribution of the variable(s) of interest among sub-units.

Key Words: *cluster sampling, detection probability, India, Markov model, occupancy modeling,* Panthera tigris, *spatial dependence, spatial replication, tigers, trap response model*

Introduction

Large-scale occupancy surveys and monitoring programs are now used throughout the world to inform conservation decisions. Most occupancy modeling requires some sort of replication in order to obtain the information needed to estimate detection probability and, thereby, probabilistically separate true absence from presence and nondetection (MacKenzie et al. 2006, Royle and Dorazio 2008). The usual sampling situation involves multiple visits to each sample unit during some period of time over which the units are assumed to be closed to changes in true occupancy. In some cases it is possible to substitute spatial replication for temporal replication in order to obtain this information about detection probability. Specifically, multiple survey sites or locations are selected from each sample unit randomly and with replacement and are then surveyed a single time, frequently on the same day (MacKenzie et al. 2006). Such a design permits estimation of occupancy at the level of the sample unit (not at the level of the specific sites or locations within each unit). When the species (or sign of the species) occupies a sampling unit, but is not present at all sites within the sampling unit, detection probability consists of two components; 1) Pr(present at survey site), and 2) Pr(detection | present at survey site). Resulting estimates of detection probability in such designs correspond to each specific survey site and are reasonable estimates of the product of these components.

Sometimes, surveys employ spatial replication but do not follow the recommendation to sample randomly and with replacement. Such survey designs typically result from logistical constraints and represent a compromise between statistical requirements and practical requirements of field surveys. It is unclear how deviations from random spatial sampling with replacement affect estimates of occupancy. In this paper, we focus on a specific field design that is being used in a large-scale occupancy survey of tigers (*Panthera tigris*) in southwestern India (Karanth and Kumar 2008, unpublished report). This type of survey design has been adopted elsewhere in southeast Asia by several conservation organizations for use with carnivores and elephants, is currently being used for carnivore surveys in Africa and is being considered for surveys of carnivores in the new world as well. The design uses spatial replication without the requisite random selection and replacement and is suspected to produce detection data that are correlated on adjacent sample sites (replicates). We first develop a new model that deals explicitly with data resulting from this type of spatially replicated design. We also develop a new "trap response" occupancy model (see MacKenzie et al. 2006) for possible use with such data. We then use computer simulation to investigate the consequences of this kind of spatial design for estimates based on (1) standard occupancy models that ignore the spatial dependence, (2) the new trap response model, and (3) the new spatial Markov process model developed for this design. Finally, we apply all three classes of model to initial data from a field survey of tigers in India (Karanth and Kumar 2008, unpublished report), focusing on variation among resulting estimates and on the results of model selection.

Survey Designs with Correlated Spatial Replication

Field Surveys

The field survey that motivated this work was developed for tigers in the 22,000 km^2 Malnad-Mysore Tiger Landscape in Karnataka State, southwestern India. Details of this specific survey will be provided below. This and similar designs designate as sampling units large geographic grid cells at a scale appropriate to the study organism, depending on biology of the species. For example, for tigers in southwestern India, the cell size was set based on expected maximum home range sizes of ~ 200 km^2 (Karanth and Sunquist 2000). Tigers are known to use forest roads and trails as travel routes and to mark them intensively with tracks, scent and scats (Karanth and Sunquist 2000). Search for tiger signs along such trails increases detection probabilities far above those expected based on random sampling. Within each cell the design ensured that each survey team was made to pass through a point randomly chosen before the survey. The trail is then walked by three trained surveyors looking for sign of the species of interest (e.g., tiger scat and tracks). The distance covered is subdivided into segments of equal length (e.g., 1 km) that are then treated as geographic replicates in the occupancy analysis. We initially viewed the chief virtue of this design as logistical feasibility (but see discussion).

Recommendations for use of geographic replicates in occupancy surveys include sampling randomly selected replicates with replacement (MacKenzie et al. 2006).

Logistical considerations sometimes cause biologists to select designs, such as that used in the tiger survey, that do not sample the replicates (segments) with replacement. The reasons for preferring sampling with replacement are not likely to be important for the large, wide-ranging, highly mobile species to which this type of survey design is applied. High mobility insures that all possible detection histories have some non-negligible *a priori* probability of occurrence. This is not the case for spatial replication of sedentary organisms that either inhabit a replicate or not. For such species, if replicate 1 is inhabited and replicate 2 is not, then only detection histories 10 and 00 are possible if sampling is without replacement. This latter situation leads to biased estimates of occupancy and detection probability. Despite the random selection of one of the replicates (e.g., trail segments) in the carnivore survey design, all other replicates are not selected at random (Figure 1). Trails are selected for surveys because they represent likely travel routes for study species. Because individual animals walk along trails for distances that can exceed segment length, detection of sign on one segment likely will translate to an increased probability of detecting sign on the next segment. In this paper, we focus on methods for dealing with data from designs with this type of potential for spatial correlation in detection probability.

This type of sampling design can be thought of generally as a cluster sampling design (e.g., Thompson 2002). Formal inference for such a design requires a model of within cluster variability (i.e., a description of the process generating variation among sub-samples). In the present context, we require a model for variability in species presence among sub-samples. The basic null model which has justified previous applications of spatial sub-sampling (e.g., see Williams et al. 2002: 555-573) is that sub-sample occupancy states are independent and identically distributed Bernoulli outcomes, in which case sub-sample occupancy becomes confounded with detection probability (as noted in paragraph 1; also see Cam et al. 2002). Random sampling with replacement is simply an attempt to induce this null model. In cluster sampling designs such as that described above, the sub-samples are spatially organized, suggesting certain types of models that accommodate spatial dependence. When such models are reasonable, they provide information that can be used to estimate additional parameters describing variation among spatial sub-samples within a cluster.

New Markov Process for Segment Occupancy, Model $[\psi(.), \theta(.), \theta'(.), p(.)]$

Under the above design, we suspected some degree of positive correlation between detection events on adjacent sample segments. Specifically, we considered the biologically likely scenario that an individual tiger might travel along a trail for several adjacent 1-km segments. We chose to model this scenario by decomposing the detection process into 2 components, (1) animal presence at a segment and (2) animal detection, given presence on a segment. We then modeled the component associated with animal presence as a 1st order spatial Markov process (e.g., Williams et al. 2002:197-202). Initially, we believed that this model might be so general that parameters would not be identifiable. We thus developed an additional model patterned after the trap response models of capture-recapture. Although we found that inference is indeed possible for the more general model, we present the new trap response model in Appendix A, as we believe that it may be useful in some sampling situations as well.

We assume that we begin the survey at one end of the survey route (trail) for a sample unit and proceed in order along the K total segments of the route. The data resulting from such a survey are detection histories (e.g., based on tiger sign) for each of the s sample units (large cells) selected to survey. Each detection history is simply a vector containing a “0” or a “1” for each of the K elements (segments), depending on whether the species was (1) or was not (0) detected on that segment. For example, consider the detection history for cell j, $h_j = 01011$. There were no detections of tiger sign on segments 1 or 3, but sign was detected on segments 2, 4 and 5.

Define the following parameters of a spatial process model for detection history data obtained under the above design (we drop the subscript j denoting the sample unit):

p = Pr (detection at a segment | sample unit occupied and species present on segment);

ψ = Pr (sample unit occupied);

θ = Pr (species present on segment | sample unit occupied and species not present on previous segment);

θ' = Pr (species present on segment | sample unit occupied and species present on previous segment).

These parameters can be used to develop a Markov spatial process model that involves an unobservable random variable, whether the species was present or not in the previous segment. The dot notation [dot following a model parameter symbol, e.g., $\psi(.)$], in the model description indicates no variation from segment to segment in a model parameter.

Consider the detection history, $h_j = 01011$. The probability associated with this history under the Markov spatial process model is:

$$\Pr\,(h_j = 01011) = \psi[(1-\theta)\theta + \theta(1-p)\theta']p[(1-\theta')\theta + \theta'(1-p)\theta']p\theta'p \quad .$$

The initial ψ in the above expression corresponds to the event that the species was present in the sample unit. The first set of brackets contains two additive components dealing with the uncertainty about segment 1. One possibility is that the species was present on segment 2 (the probability corresponding to this event is θ), but absent from the previous segment (1- θ). The other possibility is that the species was present on segment 2 (θ'), and present but undetected [θ (1-p)] on the previous segment. The p following these bracketed terms indicates detection on segment 2. The second set of brackets contains probabilities associated with the two possible outcomes on segment 3, absence of the species and presence with nondetection. There is no ambiguity associated with segments 4 and 5 (the species was present and detected at each of these segments) so the modeling is simpler.

The probability associated with a detection history of all 0's must include the probability of true absence from surveyed segments and the probability of presence and nondetection. The latter probability in turn depends on the realization of the spatial Markov process that dictates true presence and absence among segments. Consider a simple design with only 2 segments (not recommended, but this produces a probability statement that is relatively easy to follow). The probability associated with history h_j = 00 can be written as:

$$\Pr\,(h_j = 00) = \quad (1-\psi) + \psi\left[\theta(1-p)(1-\theta'p) + (1-\theta)(1-\theta p)\right]$$

The first term, (1-ψ), corresponds to the probability that the sample unit is not occupied. If the sample unit is occupied, then the first of the two main additive terms within the brackets corresponds to the probability that the first segment is occupied, but the species not detected. The next segment is then either occupied with no detection [$\theta'(1-p)$] or not occupied ($1-\theta'$). Note that $1-\theta' p = \theta'(1-p)+(1-\theta')$. The second main additive term within the large brackets corresponds to the probability that the first segment is not occupied. The second segment then may be occupied with no detection or not occupied.

Every detection history can be modeled in this manner and the likelihood under this model can be obtained simply as the product of the probabilities corresponding to all detection histories:

$$L(\psi,\theta,\theta',p,|\,h_1,h_2,...h_s)=\prod_{j=1}^{s}\Pr(h_j).$$

A general computing expression for Pr(h_j) is provided in Appendix B.

Estimates under this model can be obtained via maximum likelihood, and the senior author (JEH) has incorporated this model into program PRESENCE (Hines 2006). This software can be used to obtain estimates under this and related models. In fact, the model structure incorporated into PRESENCE is more general than that described above in that it deals with missing observations. The above model can also be implemented using a Markov chain Monte Carlo (McMC) approach. A brief description of this approach, together with WinBUGS code for this model, is presented in Supplement 1. Results of a small simulation study comparing results of the likelihood-based and McMC approaches are provided in Appendix C.

Methods

Simulation Study

A simulation study was conducted to evaluate the performance of standard occupancy models and the two new models described above in the face of data collected under the described spatial cluster sampling. The simulation of detection history data proceeded in a straightforward manner and followed the development of model $[\psi(.), \theta(.), \theta'(.), p(.)]$. For each sample unit, occupancy was determined as a Bernoulli trial with probability ψ. For those units that were occupied, presence of the species on the first segment of the survey route was determined as a Bernoulli trial with probability θ. If the species was determined to be present on segment 1, then its presence on segment 2 was determined as a Bernoulli random variable, Bern (θ'). If the species was not present on segment 1, then presence on segment 2 was determined as Bern (θ). Species presence and absence at the remaining segments were determined similarly, with probability of presence at segment *t* depending on presence or not at segment *t*-1. Once presence-absence was determined for all *K* segments of the survey route, the detection process was simulated. At each segment at which the species was present, detection was determined as Bern (*p*). This procedure led to detection histories for all *s* sample units.

Specific values were assigned to the above parameters, and detection data were generated. These detection histories were then used in conjunction with four specific occupancy models in order to assess estimator performance in the face of the described spatial sampling with correlation. Specifically, we simulated data from a study of 200 sample units, each sampled by a “trail” consisting of 10 segments. Occupancy at the level of the sample unit was set at $\psi = 0.75$. For occupied sample units, probability of animal presence for segment 1 and for all subsequent segments, *t*, for which there was no presence on the previous segment, *t*-1, was set at $\theta = 0.1$. For occupied sample units, probability of animal presence on any segment, *t*, for which there was presence on the previous segment, *t*-1, was set at $\theta' = 0.5$. The detection probability for each segment, conditional on presence of animals on the segment, was set at $p_t = p = 0.80$. The detection data were thus generated according to a spatial Markov process and showed a strong positive spatial correlation between successive segments.

In order to provide a set of simulations to serve as a sort of control, we also generated data that corresponded to standard occupancy model assumptions (MacKenzie et al. 2002, 2006; Royle and Dorazio 2008). These simulations were carried out with the same sample sizes and parameter values as above, with the exception that we set $\theta = \theta' = 0.3$. This latter constraint corresponds to the situation in which occupancy of a particular segment is the same regardless of whether the previous segment is occupied (we have removed the Markovian dependence).

These values were used to generate 1000 sets of detection history data corresponding to each of the above spatial processes (Markovian and non-Markovian). These data were then used to estimate occupancy and detection probability under two standard occupancy models that were not designed to account for the possibility of spatial correlation between segments, $[\psi(.), p(.)]$ and $[\psi(.), p(t)]$. The first model assumes constant occupancy among sites and constant detection probability among sites and segments. Note that this model corresponds to the second set of simulations described above. The second model assumes constant occupancy among sites and constant detection probability among sites, but permits variation in detection probability among segments. We then evaluated the performance of estimators from the new model developed specifically for such data, $[\psi(.), \theta(.), \theta'(.), p(.)]$. For comparison, we also fit the new trap response model [$\psi(.), p(.), p'(.)$] developed as an approximation to the process generating the data (Appendix A).

In evaluating the performance of all of these models, we focused on the bias of the occupancy estimator, $\hat{\psi}$, and computed bias as:

$$Bias(\hat{\psi}) = \frac{\sum_{k=1}^{n}(\hat{\psi}_k - \psi)}{n} ,$$

where $\hat{\psi}_k$ is the estimate obtained for simulated data set *k*, *n* is the number of simulations ($n = 1000$), and ψ is the true occupancy value used to generate the data. Relative bias is obtained by dividing bias by the true parameter value, e.g.,

$$RelBias\,(\hat{\psi}) = Bias(\hat{\psi}) / \psi .$$

We also evaluated the performance of the estimates of standard errors for various estimators, $\hat{\phi}$. We computed bias of these estimators as:

$$Bias[S\hat{E}(\hat{\phi})] \approx \frac{\sum_{k=1}^{n} S\hat{E}(\hat{\phi}_k)}{n} - \sqrt{\frac{\sum_{k=1}^{n}(\hat{\phi}_k - \hat{\bar{\phi}})^2}{n-1}} ,$$

basically computing the difference between the average model-based standard error and the iteration-based estimate based on the 1000 estimates of ϕ.

Large-scale Tiger Occupancy Survey

The spatial distribution survey of tigers in Karnataka State, India (Karanth and Kumar 2008; unpublished report) was conducted between February 2006 and June 2007 across a 22,000 km^2 area, assuming that tiger spatial distribution remained unchanged during this relatively short period. Based on previous studies of tiger density, home range size and relationship to abundance of ungulate prey (Karanth and Sunquist 2000, Karanth et al. 2004), we assumed an expected maximum home-range size of 150 km^2 for tigers. The grid cell size selected was larger than this area, primarily with the goal of eventually linking the occupancy parameter to a measure of tiger abundance using the Royle-Nichols (2003) model.

The survey was carried out over 205 grid cells of 188 km^2 each that coincided with the survey map-grid feature to facilitate field work. The surveyed distance walked within each cell was 40 km if the cell entirely comprised tiger habitat based on land-cover features. This distance was proportionately reduced depending on extent of habitat, and cells with less < 10% forest cover were not surveyed, as they were unlikely to shelter tigers. Thus, the number of 1-km replicate segments surveyed per cell ranged from 4 to 42. The tiger signs (tracks, scats) encountered were verified and recorded (Karanth and Kumar 2008, unpublished report).

As noted above, within each sample unit (grid cell) the design ensured that each survey team passed through a point randomly chosen before the survey. However, the starting point of the survey was not necessarily located at the beginning of the trail. The first segment surveyed was typically in the interior of the trail. This design requires a slight modification of the probability structure described above for model $\psi(.), \theta(.), \theta'(.), p(.)$. Specifically, it is not appropriate to model the probability of segment occupancy for the initial segment as θ, because the initial surveyed segment may or may not be preceded by an occupied segment. Instead, we need an expression for the probability that a randomly selected segment from the interior of a trail is occupied. A reasonable expression for this probability is the equilibrium probability of occupancy for a spatial Markov process defined by θ and θ', as given by:

$$\frac{\theta}{\theta + (1 - \theta')} \tag{1}$$

(e.g., see MacKenzie et al. 2006:208). Thus, the probability associated with segment-level occupancy of the initial surveyed segment can be modeled either as θ, when this segment is at the beginning of a trail, or as expression (1) for surveys such as the tiger survey in which the initial surveyed segment is in the trail interior.

We fit six models to the tiger data set. The first three models were parameterized with constant occupancy and three different models of the detection process, $[\psi(.), \theta(.), \theta'(.), p(.)]$, $[\psi(.), p(.), p'(.)]$, and $[\psi(.), p(.)]$. The additional models reflected the same three models of the detection process, but with occupancy modeled as a linear-logistic function of the number of segments surveyed (because this number reflected the proportion of the cell comprised of tiger habitat, denoted h in model notation), $[\psi(h), \theta(.), \theta'(.), p(.)]$, $[\psi(h), p(.), p'(.)]$, and $[\psi(h), p(.)]$. All models were implemented in program PRESENCE (Hines 2006), and maximum likelihood estimates were computed. AIC was computed as a model selection statistic, and AIC weights were computed for the six models (Burnham and Anderson 2002). We considered use of $\mathrm{AIC_c}$, AIC adjusted for small sample size. However, the issue of just what constitutes sample size in the case of occupancy models has not been resolved, so we opted for the unmodified AIC. Note that the issue of defining sample size is problematic not only for occupancy modeling, but for many other situations as well (Burnham and Anderson 2002:332-333).

For a given model of occupancy (e.g., constant), we expected the model with spatial segment-level dependency, $[\psi(.), \theta(.), \theta'(.), p(.)]$, to be selected as the most appropriate model for the data. This model was developed for exactly this kind of spatial process. Simulation study results led us to suspect that the trap response model,$[\psi(.), p(.), p'(.)]$, would provide a fair description of the data, and that the standard occupancy model, $[\psi(.), p(.)]$, would neither describe the data well nor provide a good estimate of occupancy. For a given model of the detection process (e.g., constant), we expected the model with occupancy written as a function of the habitat covariate, $[\psi(h), p(.)]$, to perform better than the constant occupancy model, $[\psi(.), p(.)]$. Overall, we thus expected model $[\psi(h), \theta(.), \theta'(.), p(.)]$ to perform best and model $[\psi(.), p(.)]$ to perform worst.

Results

Simulations

When data were simulated under the case of no spatial correlation among segments (Table 1, column $\theta = \theta' = 0.3$), parameter estimates for the two standard occupancy models, $[\psi(.), p(.)]$ and $[\psi(.), p(t)]$, were very close to expectations. Average occupancy estimates were only slightly larger than the true value of 0.75; bias and relative bias were small. True segment-level detection probability for these standard models can be obtained as the product of segment-level occupancy and detection, conditional on occupancy, $\theta p = 0.3 * 0.8 = 0.24$. The average values of $\hat{p}$ under these 2 models were again nearly identical to 0.24 (Table 1). Similarly, the model-based estimates of standard error performed well, as they matched the empirical estimates of standard deviation of parameter estimates quite closely (Table 1).

However, when data were simulated with spatial correlation (Table 1, column $\theta = 0.1, \theta' = 0.5$) the standard models performed very poorly. For example, relative bias of the occupancy estimators under these standard models was about -0.30, indicating that occupancy estimates were about 30% too small. The spatial Markov process induced heterogeneity among segments with respect to segment-level occupancy, with segments preceded by an occupied segment having very different probabilities of being occupied than segments not preceded by an occupied segment.

We recognized that the Markov detection process (trap response) model, [$\psi(.), p(.), p'(.)$], did not capture the underlying model of spatial dependence perfectly, but we hoped that it would provide an approximation that yielded reasonable estimates. The occupancy estimator under this model performed much better than the standard occupancy models, but still exhibited negative bias (relative bias nearly -0.07). The estimated detection probability for segments preceded by segments with a detection ($\hat{p}'$) should estimate the product, $\theta' p = 0.5 * 0.8 = 0.4$, and the average of the estimates was very close to this value (Table 1). However, detection probability for segments not preceded by segments with a detection (*p*) pertain to: (1) some segments for which preceding segments were unoccupied, and (2) other segments for which previous segments were occupied yet not detected. It is this detection parameter which is not estimated properly, leading to the negative bias in the occupancy estimator.

The model [$\psi(.), \theta(.), \theta'(.), p(.)$] that was developed explicitly for the underlying spatial model used to generate the data performed well with small positive bias in the occupancy estimator (relative bias less than 0.02). Our main question about implementing this model was whether the underlying parameters were identifiable. The model was not as numerically stable as the other 3 models with convergence failure occurring in about 150 of the 1000 simulations. Because this was a simulation study, there was no attempt to use alternative starting values or try other approaches to obtaining convergence in problem cases. When this model was fit to data generated with no spatial dependence ($\theta = \theta' = 0.3$), it produced reasonable estimates of occupancy and its variance (Table 1). However, the absence of Markovian spatial dependence in the process used to generate the data produced difficulties in estimating the parameters associated with the detection process (the local occupancy and detection parameters) and also led to increased numerical instability (convergence failure in nearly ¼ of the simulations).

Tiger Survey

As expected, model selection results for the tiger survey data strongly support the use of the Markov process models for spatial dependence, [$\psi(h), \theta(.), \theta'(.), p(.)$] and [$\psi(.), \theta(.), \theta'(.), p(.)$] (Table 2). The former model including the effects of habitat on occupancy was favored, as expected, and received a model weight of about 0.87. The coefficient associated with proportion of the cell in tiger habitat was estimated to be $\hat{\beta}_1 = 0.053, S\hat{E}(\hat{\beta}_1) = 0.021$. Thus, under the selected model, $\psi(h), \theta(.), \theta'(.), p(.)$, the probability of occupancy at the level of the 200 km^2 cell increased with the proportion of habitat in the cell as predicted. The occupancy estimate for a cell with the average proportion of habitat was $\hat{\psi}(\bar{h}) = 0.50, S\hat{E}[\hat{\psi}(\bar{h})] = 0.069$. The other model that received support, [$\psi(.), \theta(.), \theta'(.), p(.)$], yielded an occupancy estimate of $\hat{\psi} = 0.57, S\hat{E}(\hat{\psi}) = 0.077$ (Table 3). The trap response models, [$\psi(h), p(.), p'(.)$] and [$\psi(.), p(.), p'(.)$], developed as crude approximations to the true detection process, received virtually no support, but still provided a relatively better description of the data than did the standard occupancy models, [$\psi(h), p(.)$] and [$\psi(.), p(.)$]. As predicted based on simulation results, the trap response and standard occupancy models yielded progressively smaller estimates of occupancy (Table 3). The naïve estimate computed as the proportion of cells at which tiger sign was detected was 0.36.

The parameter estimates corresponding to the detection process provided strong evidence of the kind of process for which the model was developed. Under the top model, the segment level occupancy for segments not preceded by an occupied segment was only $\hat{\theta} = 0.07$, whereas occupancy for a segment preceded by an occupied segment was estimated to be $\hat{\theta}' = 0.79$. The segment-level detection probability, conditional on segment-level occupancy, was estimated to be relatively high as expected, $\hat{p} = 0.42$. This value is larger than the detection parameter estimates under the other two classes of models (Table 3), because these latter parameters incorporate both segment-level occupancy and detection given occupancy.

Discussion

This work was designed to investigate ways of estimating occupancy using spatial replication in a case where replicate-level occupancy follows a 1-dimensional spatial

Markov process. We described the basic kind of field survey design that motivated this work and noted that this design is frequently used for surveying large, wide-ranging mammals that typically use trails for marking and movement. We believe this approach has great potential utility because it explicitly models the process that generates signs along trails (animal behavior) and also meets key logistical and practical needs of field surveys in tropical forests (easy movement of survey teams across rugged landscapes using trails and ease of detecting animal signs on them). In this design spatial replicates are visited in a specified order (e.g., segments along a trail), and we suspected that replicate-level occupancy was correlated for adjacent segments (because of behavior of tigers, Karanth and Sunquist 2000). We developed a new occupancy model to deal explicitly with this sort of spatial process. We also developed a new trap response occupancy model as a rough approximation for data obtained under such sampling designs. We then conducted a simulation study generating detection history data according to the hypothesized spatial model for such a survey in order to evaluate estimators obtained under standard occupancy models (MacKenzie et al. 2006) and the two new models that we developed.

In the case of higher probability of local occupancy of a trail segment given local occupancy of the preceding trail segment, occupancy estimators for standard models showed substantial negative bias. The new trap response occupancy model $[\psi(.), p(.), p'(.)]$ simply places a Markovian dependence on detection probabilities, such that detection probability for a spatial replicate (trail segment) depends on whether or not sign of the species was detected on the preceding replicate or trail segment. We were confident that such a model could be fit to detection history data and speculated that it might provide a reasonable approximation to the hypothesized underlying spatial process. Simulation results indicated that occupancy estimates from this model were indeed much improved over those of the standard occupancy models, but they were still negatively biased.

The other new occupancy model [$\psi(.), \theta(.), \theta'(.), p(.)$] was developed specifically for the Markov spatial process that was thought to characterize the tiger survey data and that was used to generate the simulated data. However, we were uncertain about whether this model could be readily fit to detection history data and whether the model parameters were really identifiable. The model failed to converge for about 15% of the simulated data sets, but we made no efforts to use different starting values or otherwise deal with the convergence problems. For the remainder of the data sets, convergence was obtained and model estimators performed very well, exhibiting negligible bias.

We fit six models to the data from the large-scale occupancy field survey for tigers in southwestern India that motivated the simulation study. We modeled occupancy as either a constant or a function of the proportion of the grid cell that contained suitable tiger habitat. For each type of occupancy model, we modeled the detection process either using a constant detection probability, the new trap response model, or the new spatial Markov process model. The AIC model selection statistics provided support for the habitat model of occupancy and strong support for the Markov spatial process model, as predicted. The parameter estimates associated with segment-level occupancy provided evidence of a large increase in the probability of local occupancy of a segment when the preceding segment was occupied. The trap response models specifying a Markov process for detections indicated a much higher segment-level detection probability for segments preceded by a detection. The results of these models were consistent with our predictions based on tiger behavior and field survey methods.

Results of the field survey analysis provided evidence that tiger biologists are capable of identifying habitat elements of the Malenad–Mysore Tiger Landscape of Karnataka that are essential for tiger existence and strengthen their arguments for protection of the areas of suitable habitat that do remain. Cells containing large amounts of habitat showed the largest probabilities of tiger occupancy, providing stronger inferences than expert statements about specific areas most important to tigers. The estimated fraction of surveyed cells that were occupied under the most appropriate model was about 0.50, whereas the naïve occupancy estimate obtained as the proportion of cells at which tigers were detected was 0.36. The Karnataka survey demonstrates the ability to assess current range of secretive animals using large-scale field surveys, and the analytic results indicate the need to adequately deal with nondetection in analysis of resulting data. Karanth et al. (in preparation) are currently investigating specific factors within each cell, in addition to available habitat, that may influence probability of a cell being occupied by tigers.

The occupancy estimates under the six models were also consistent with expectations for a Markov spatial process based on simulation results. The standard occupancy models yielded the smallest occupancy estimates, whereas the Markov spatial process models produced the largest occupancy estimates ($\hat{\psi}(\bar{h}) = 0.50, \hat{\psi} = 0.57$).We conclude that the spatial process models were appropriate for the data and provided an occupancy estimate that was a substantial improvement over the naïve estimate.

We have not extended the kind of modeling presented here to the case of dealing with abundance-induced heterogeneity and abundance estimation (e.g., Royle and Nichols 2003). At present, those interested in such models (or indeed in any other occupancy models requiring independent spatial replicates) can use the model presented here to test for spatial dependence at different segment lengths. For example, the tiger survey data were collected at 1 km intervals, but they can be aggregated to create longer segments of 2 km, 3 km, 4 km, etc. Spatial dependence is expected to decline with segment length, and the model presented here can be used to test for spatial independence in order to select a length at which independence is a reasonable approximation (Karanth and Kumar 2008, unpublished report). Models requiring independence can then be used with these aggregated data sets to draw inferences.

The issue of being able to break trails into segments of various lengths poses the interesting design question of whether there is an optimal segment length with respect to precision or mean squared error. It would be unwise to use segment lengths so short that most of them contain no detections even when animals are present at the level of the sample unit. Similarly, segments should not be so long that <3 segments exist in most sample units. Beyond those simple recommendations, it does not seem possible to provide guidelines, or even construct a simulation study, that would be generally useful. Instead, this is the sort of design issue that should be investigated (e.g., via simulation) on a case by case basis, and we suspect that the conclusion will depend very heavily on the logistics and biology of the survey situation.

Both new models were developed specifically for use of spatial replicates for occupancy estimation in the face of Markov spatial processes in detection or replicate-level occupancy. In addition to the increasing use of the described type of survey design for large mammals worldwide, we also foresee application to any road- or trail-based surveys (track surveys, scent station surveys) of species that sometimes use roads and trails as travel routes. Even for species that do not themselves use roads or trails as travel routes, we believe that the models developed here may prove useful. For example, the North American Breeding Bird Survey (e.g., Peterjohn and Sauer 1993) is a road-based survey using avian point counts conducted at 50 stops located along the route at 0.8 km intervals. Adjacent stops are expected to exhibit greater similarity of habitat, on average, than stops located farther apart. In some instances, individual birds with range centers located between two adjacent stops may be detected at each of two stops. Both of these possibilities would be expected to generate the sort of Markovian spatial dependence for which our models were developed.

In addition to applications for certain spatial sampling designs, we believe that these models will also have broad application to occupancy studies that use temporal replication. Consider an occupancy study of a territorial species such that only a single individual or pair or group (e.g., a wolf pack) is likely to use any particular sampling unit during the survey season. Further assume that individuals of the species travel widely, such that a sample unit is likely to be used by the species on some days and not on others. If this use is correlated in time (e.g., a tiger or a wolf pack cycling through a large range, perhaps spending multiple days at a kill site and then moving on), then either of the proposed new models might be a reasonable candidate for describing that situation.

Other situations for which temporal Markov processes may be useful involve strong seasonality of animal presence and use of surveyed sites. For example breeding anuran occupancy can be modeled as a function of date (season) and air temperature, with substantial pulses of breeding activity (Weir et al. 2005). Markovian modeling, with initial local occupancy (θ) a function of such environmental covariates and subsequent local occupancy (θ') probability larger following initial emergence, might be useful in such a situation. Occupancy analyses of butterfly surveys may show multiple peaks of emergence within a season (Kery et al. 2009), and Markovian models for such data should be useful as well. In such situations the suggested Markovian modeling should not only permit reasonable inference about sample unit occupancy, but also permit inference about the temporal pattern of local occupancy (see Kery et al. 2009 and discussion below). In the case of temporal Markovian dependence, estimates of θ and θ' pertain to an occupancy process analogous to temporary emigration of individual animals in capture-recapture studies (e.g., Kendall et al. 1997).

When viewed in the broader context of models for estimation of animal abundance and occupancy, model [$\psi(.), \theta(.), \theta'(.), p(.)$] represents an explicit decomposition of the detection process into two components: availability and detection given availability (see discussion in Nichols et al. 2008*b*). With respect to abundance estimation, availability refers to an individual animal having a non-zero probability of detection, for example because it is on the water surface, not submerged (e.g., Marsh and Sinclair 1989), at the time of a visual survey, or because it vocalizes at the time of an auditory survey (e.g., Farnsworth et al. 2002), or because it is present in the sampled area (not a temporary emigrant) during a capture period in a trapping study (Kendall et al. 1997). In cases of occupancy estimation for mobile animals that use areas larger than a single sample unit, the detection process can also be decomposed into two components:

(1) local presence of species at the temporal or spatial replicate and (2) detection given local presence. Similar decomposition was also used to estimate occupancy at two different spatial scales using data from multiple detection devices at the local sampling site (Nichols et al. 2008*a*). This sort of decomposition of the detection process may be needed to properly model detection, as in the present study, and it may also provide the ability to address interesting ecological questions (Nichols et al. 2008*a*).

The kind of modeling used here may also be relevant to the still broader application area of cluster sampling (e.g., Thompson 2002). Specifically, the process and/or observation model for the variable of interest (e.g., occupancy) might differ between the two levels at which sampling occurs, the cluster and the sub-unit within a cluster. It is possible that there are other situations in which explicit modeling of the process governing the distribution of the variable across sub-units might yield improved inference. In the spirit of placing this modeling within a general context, we also note that this work represents a special application of inference for hidden Markov chains (e.g., Cappé et al. 2005).

We note the potential to extend the type of modeling presented here to capture-recapture modeling of Markovian temporary emigration in demographically “closed” animal populations. Kendall et al. (1997) described approaches to the modeling of temporary emigration for open populations using Pollock’s (1982) robust design, and they included Markovian models. Kendall (1999) showed that abundance estimates under capture-recapture models for closed populations are robust to random (non-Markovian) temporary emigration. However, Markovian temporary emigration, such that presence of an animal on a trapping array at one sample period depends on whether it was present or not the previous sample period, was shown to induce bias in abundance estimates (Kendall 1999). The type of Markovian modeling used here can be used in closed capture-recapture modeling to permit estimation of abundance in the face of such Markovian temporary emigration.

We conclude with a consideration of the two spatial sampling designs that motivated this work (Figure 1). We introduced the topic by noting that use of spatial replicates with standard occupancy models strictly requires that spatial replicates be selected randomly and with replacement from each sample unit (e.g., grid cell). Kendall and White (in review) provide some results on the magnitudes of bias that can arise during certain sampling situations when sampling is not conducted with replacement. In the absence of spatial dependence of segment-level occupancy (this is the spatial model that we attempt to induce by random sampling), standard occupancy models appear to perform reasonably (results presented above). However, in the presence of spatial dependence, standard models yield biased estimates of grid cell occupancy. The model that we present here can be used to test for such dependence, and if it is present, to provide reasonable parameter estimates. Our initial development of the Markov spatial model was motivated by biological and logistical considerations associated with certain kinds of field surveys. However, the Markovian dependence that we initially viewed as a nuisance actually permits decomposition of the detection process in a manner that is not possible with independent segments (e.g., see the poor estimator performance for $\hat{\theta}$, $\hat{\theta}'$, and $\hat{p}$ in last column of Table 1 for model [$\psi(.), \theta(.), \theta'(.), p(.)$]). Thus, if we are interested either in decomposing the detection process, or in focusing for other reasons on lower-level occupancy, then sampling designs that lead to Markovian spatial dependence provide an advantage.

Acknowledgments

We thank Shannon Barber-Meyer and Sunarto Sunarto for asking questions about consequences of trail-based sampling designs for occupancy estimation. Two anonymous referees provided constructive comments on the initial submission that were useful in preparing this revision.

Literature Cited

Burnham, K.P., and D.R. Anderson. 2002. Model selection and multimodel inference. 2nd edition. Springer, New York.

Cam, E., J.D. Nichols, J.E. Hines, and J.R. Sauer. 2002. On the estimation of species richness from species accumulation data. Ecography 25: 102-108.

Cappé, O., E. Moulines, and T. Rydén. 2005. Inference in hidden Markov models. Springer, New York.

Farnsworth, G.L., K.H. Pollock, J.D. Nichols, T.R. Simons, J.E. Hines, and J.R. Sauer. 2002. A removal model for estimating detection probabilities from point count surveys. Auk 119:414-425.

Hines, J. E. 2006. PRESENCE2 - Software to estimate patch occupancy and related parameters. USGS-PWRC. http://www.mbr-pwrc.usgs.gov/software/presence.html.

Karanth K. U., Kumar N. S., Gopalaswamy A., and Srinivas V. 2008. Distribution and dynamics of tiger and prey populations in Karnataka. Wildlife Conservation Society-India Program, Centre for Wildlife Studies (unpublished report).

Karanth, K. U**.,** Nichols, J. D., Kumar, N. S., Link, W. A. and Hines, J. E. 2004. Tigers and their prey: predicting carnivore densities from prey abundance. Proceedings of the National Academy of Sciences, USA 101: 4854-4858.

Karanth K. U. and Sunquist M. E. 2000.. Behavioural correlates of predation by tiger, leopard and dhole in Nagarahole, India. Journal of Zoology, London 250: 255-265.

Kendall, W.L. 1999. Robustness of closed capture-recapture methods to violations of the closure assumption. Ecology 80:2517-2525.

Kendall, W.L., J.D. Nichols and J.E. Hines. 1997. Estimating temporary emigration and breeding proportions using capture-recapture data with Pollock's robust design. Ecology 78:563-578.

Kery, M. , J.A. Royle, M. Plattner, and R.M. Dorazio. Species richness and occupancy estimation in communities subject to temporary emigration. Ecology 90:1279-1290.

MacKenzie, D.I., J.D. Nichols, G.B. Lachman, S. Droege, J.A. Royle, and C.A. Langtimm. 2002. Estimating site occupancy when detection probabilities are less than one. Ecology 83:2248-2255.

MacKenzie, D.I., J.D. Nichols, J.A. Royle, K.H. Pollock, L.A. Bailey, and J.E. Hines. 2006. Occupancy modeling and estimation. Academic Press, San Diego, CA. 324pp.

Marsh, H. and D.F. Sinclair. 1989. Correcting for visibility bias in strip transect aerial surveys of aquatic fauna. Journal of Wildlife Management 53:1017-1024.

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

Nichols, J.D., L. Thomas, and P.B. Conn. 2008*b*. Inferences about landbird abundance from count data: recent advances and future directions. Pages 201-235 *in* D.L. Thomson, E.G. Cooch, and M.J. Conroy, eds. Modeling demographic processes in marked populations. Springer, NY.

Peterjohn, B. G., and J.R. Sauer. 1993. North American breeding bird survey annual summary 1990-91. Bird Populations **1**:1-15.

Pollock, K.H. 1982. A capture-recapture design robust to unequal probability of capture. Journal of Wildlife Management 46:752-756.

Royle, J.A., and R.M. Dorazio. 2008. Hierarchical modeling and inference in ecology. Academic Press, New York.

Royle, J.A., and J.D. Nichols. 2003. Estimating abundance from repeated presence absence data or point counts. Ecology 84:777-790.

Sandland, R.L., and P. Kirkwood. 1981. Estimation of survival in marked populations with possibly dependent sighting probabilities. Biometrika 68:531-541.

Thompson, S.K. 2002. Sampling. Wiley, New York.

Weir, L.A., J.A. Royle, P. Nanjappa, and R.E. Jung. 2005. Modeling anuran detection and site occupancy on North American Amphibian Monitoring program (NAAMP) routes in Maryland. Journal of Herpetology 39:627-639.

Williams, B.K., J.D. Nichols, and M.J. Conroy. 2002. Analysis and management of animal populations. Academic Press, San Diego. 817pp.

Yang, H-C., and A. Chao. 2005. Modeling animals' behavioral response by Markov chain models for capture-recapture experiments. Biometrics 61:1010-1017.

Table 1. Simulation results for 1000 sets of detection history data for 200 sites under each of two spatial correlation scenarios. Detection history data were generated for a true site occupancy of $\psi = 0.75$ and segment-level detection probability of $p = 0.8$. Detection history data were generated both with ($\theta = 0.1$, $\theta' = 0.5$) and without ($\theta = \theta' = 0.3$) spatial correlation of segment-level occupancy between adjacent segments. Four different models were fit to each simulated data set, and parameter estimates under these models are summarized. Presented are the mean parameter estimates from all 1000 simulated data sets (e.g., $\bar{\hat{\psi}}$), the mean of the model-based estimates of standard error (e.g., $\bar{\widehat{SE}}(\hat{\psi})$), and the replication based estimate of standard deviation of parameter estimates ($\widehat{SD}(\hat{\psi})$). The model that included spatial correlation [$\psi(.), \theta(.), \theta'(.), p(.)$] did not always converge in the simulations, and the number of cases for which convergence was obtained is included in the table.

		True segment occupancy parameter values	
Model	**Parameter estimates**	**$\theta = 0.1$, $\theta' = 0.5$**	**$\theta = \theta' = 0.3$**
$\psi(.), p(.)$	$\bar{\hat{\psi}}$	0.5194	0.7532
	$\bar{\widehat{SE}}(\hat{\psi})$	0.0429	0.0351
	$\widehat{SD}(\hat{\psi})$	0.0448	0.0363
	$\bar{\hat{p}}$	0.1792	0.2395
	$\bar{\widehat{SE}}(\hat{p})$	0.0148	0.0124
	$\widehat{SD}(\hat{p})$	0.0189	0.0128
$\psi(.), p(t)$	$\bar{\hat{\psi}}$	0.5178	0.7515
	$\bar{\widehat{SE}}(\hat{\psi})$	0.0429	0.0352
	$\widehat{SD}(\hat{\psi})$	0.0447	0.0432
	$\bar{\hat{p}}_5$	0.1855	0.2393
	$\bar{\widehat{SE}}(\hat{p}_5)$	0.0387	0.0349
	$\widehat{SD}(\hat{p}_5)$	0.0406	0.0358

Table 1(cont.)

		True segment occupancy parameter values	
Model	**Parameter estimates**	θ **= 0.1,** θ' **= 0.5**	$\theta = \theta'$ **= 0.3**
$\psi(.), p(.), p'(.)$	$\hat{\hat{\psi}}$	0.7039	0.7534
	$S\hat{\hat{E}}(\hat{\psi})$	0.0835	0.0356
	$S\hat{D}(\hat{\psi})$	0.0894	0.0366
	$\hat{\hat{p}}$	0.0980	0.2397
	$S\hat{\hat{E}}(\hat{p})$	0.0170	0.0146
	$S\hat{D}(\hat{p})$	0.0159	0.0152
	$\hat{\hat{p}}'$	0.3977	0.2390
	$S\hat{\hat{E}}(\hat{p}')$	0.0491	0.0250
	$S\hat{D}(\hat{p}')$	0.0383	0.0226
$\psi(.), \theta(.), \theta'(.), p(.)$	$\hat{\hat{\psi}}$	0.7648	0.7534
	$S\hat{\hat{E}}(\hat{\psi})$	0.1068	0.0355
	$S\hat{D}(\hat{\psi})$	0.1054	0.0360
	$\hat{\hat{\theta}}$	0.0996	0.4843
	$S\hat{\hat{E}}(\hat{\theta})$	0.0188	0.0793
	$S\hat{D}(\hat{\theta})$	0.0188	0.2647
	$\hat{\hat{\theta}}'$	0.5082	0.4887
	$S\hat{\hat{E}}(\hat{\theta}')$	0.0686	0.0858
	$S\hat{D}(\hat{\theta}')$	0.0733	0.2670
	$\hat{\hat{p}}$	0.7955	0.6548
	$S\hat{\hat{E}}(\hat{p})$	0.1054	0.0499
	$S\hat{D}(\hat{p})$	0.1169	0.3210
	no. converged	*851*	*753*

Table 2. Model selection statistics for 6 models fit to tiger survey data from southwestern India. Statistics include ΔAIC, AIC weight, *w*, -2 times the logarithm of the likelihood, -2log(*L*), and number of parameters in the model.

Model	ΔAIC	Weight, *w*	-2log(*L*)	Parameters
$\psi(h), \theta(.), \theta'(.), p(.)$	0.00	0.87	1564.67	5
$\psi(.), \theta(.), \theta'(.), p(.)$	3.77	0.13	1570.44	4
$\psi(h), p(.), p'(.)$	25.18	0.00	1591.85	4
$\psi(.), p(.), p'(.)$	31.30	0.00	1599.97	3
$\psi(h), p(.)$	123.93	0.00	1692.60	3
$\psi(.), p(.)$	134.14	0.00	1704.81	2

Table 3. Estimates of occupancy and parameters related to the detection process for tiger survey data from southwestern India. ψ is the probability that a 188 km^2 sample unit is occupied. Definitions of p vary among models, but for the first model, it denotes the probability of detecting sign on a segment, given the presence of tigers on the segment (local occupancy). θ denotes occupancy of a segment, given that the preceding segment was unoccupied. θ' denotes occupancy of a segment, given that the preceding segment was occupied.

Model	$\hat{\psi}(S\hat{E}[\hat{\psi}])^{a}$	$\hat{p}(S\hat{E}[\hat{p}])$	$\hat{p}'(S\hat{E}[\hat{p}'])$	$\hat{\theta}(S\hat{E}[\hat{\theta}])$	$\hat{\theta}'(S\hat{E}[\hat{\theta}'])$
$\psi(h), \theta(.), \theta'(.), p(.)$	0.50(0.069)	0.42(0.063)	---	0.07(0.017)	0.79(0.061)
$\psi(.), \theta(.), \theta'(.), p(.)$	0.57(0.077)	0.42(0.059)	---	0.07(0.017)	0.80(0.057)
$\psi(h), p(.), p'(.)$	0.42(0.044)	0.09(0.008)	0.36(0.031)	---	---
$\psi(.), p(.), p'(.)$	0.46(0.045)	0.09(0.008)	0.36(0.031)	---	---
$\psi(h), p(.)$	0.38(0.039)	0.14(0.008)	---	---	---
$\psi(.), p(.)$	0.41(0.039)	0.13(0.008)	---	---	---

[a] For models in which occupancy is a function of habitat, $\psi(h)$, table shows the estimated occupancy (and *SE*) at the average value of the habitat covariate, $\hat{\psi}(\bar{h}); S\hat{E}[\hat{\psi}(\bar{h})]$. For all 3 habitat models, these estimates were within 0.01 of the average occupancy taken over all cells, $\bar{\hat{\psi}}(h)$.

Figure 1. Two sampling designs employing spatial replication to draw inference about occupancy and detection probabilities. Design 1 (left panel) depicts an example of random sampling whereas Design 2 (right panel) depicts sampling of segments along a trail, likely producing correlated spatial replicates.