Using occupancy models to improve inferences for conservation easements

Matthew A Williamson

5/22/2019

Introduction

Conservation planning is, necessarily, an activity that involves social, political, economic, and biological choices and trade-offs (Margules & Pressey, 2000). The degree-to-which the spatial juxtaposition of these factors affects the likelihood that conservation actually occurs is rarely evaluated empirically making it difficult to identify general drivers of conservation action or to develop conservation plans that account for both economic and socio-political cost (Williamson, Schwartz, & Lubell, 2018). The ability to evaluate the key factors affecting conservation adoption or identify locations likely to adopt new conservation actions requires information on where similar actions have occurred in the past. Spatially explict information on conservation actions is increasingly available, but is often incomplete due to capacity limitations or privacy restrictions (Rissman, Owley, L'Roe, Morris, & Wardropper, 2017).

Incomplete records in geospatial data on conservation adooption creates a situation analogous to that of imperfect detection often encountered wildlife population monitoring and habitat modelling. Cryptic species, variability in activity levels, and differing levels of observer skill all create situations where a species may go unobserved at a location despite being present. This can result in biased estimates of species occurrence and lead to spurious inferences with respect to the relationship between predictors and occurrence probability (Gu & Swihart, 2004, @kellner2014). Similarly, the presence of a conservation activity (e.g., a conservation easement) that goes unreported may induce similar biases in efforts to determine the predictors of adoption because locations where the activity is absent are ambiguous (i.e., the activity may truly be absent or it may be present, but unreported in the database under consideration). For example, if income increases the likelihood that an individual adopts a particular action, but decreases the likelihood that they report it, observations of that action will appear unrelated to income.

Occupancy models, a suite of flexible models that use repeated surveys to estimate the probability that a site is truly occupied by a species while accounting for imperfect detection (MacKenzie et al., 2002, 2017), are increasingly applied across a wide range of ecological problems to reduce bias induced by imperfect or heterogenous detection (reporting; Bailey, MacKenzie, & Nichols, 2014). Occupancy, defined as the proportion of sites where an object of interest (either a species or a conservation action) is present, is modeled as the result of two linked processes. The state process which describes where the object occurs and the observation process which describes how the object is detected at each sample location (MacKenzie et al., 2002; Guillera-Arroita, 2014). Although occupancy models may more accurately represent the data generation process (Guillera-Arroita, 2014), they require at least one additional axis of information relative to a more standard logistic regression or generalized linear model which may result in unstable parameter estimates and/or parameters that are not uniquely identifiable (Welsh, 2013). Further, occupancy models assume that the occupancy state is closed (i.e. occupancy does not change within the sampling season), sites are independent (detection of the object at one location does not affect detection at another), and no unexplained heterogeneity in occupancy or detectibility (MacKenzie et al., 2002). Violations of these assumptions are likely in attempts to model human behaviors as norms, social networks, and institutional arrangements may violate independence assumptions or induce unmodeled heterogeneity. Hence, occupancy models may trade one problem (i.e. unmodeled reporting probability) for another (unidentifiable or biased parameters) when applying them to conservation actions or behaviors.

Here we evaluate the utility of occupancy models for developing models of incompletely reported data on conservation actions. We use conservation easements, land-use agreements in which a landowner agrees to limit land use in exchange for direct payments or reduced tax burdens, as a motivating example becasue conservation easements are a key element of contemporary efforts to augment the existing protected areas network in the United States (Cheever & MaLaughlin, 2014) and because the location of many easments is reported in the National Conservation Easement Database (https://www.conservationeasement.us). The National Conservation Easement Database contains voluntarily reported easement boundaries and associated information for an estimated 60% of easements in the United States. We describe efforts to relate spatial predictors to easement occurrence as a situation analogous to efforts to model the occupancy probability of a species as a function of predictors describing potential elements of that species habitat. We use a series of simulation experiments to evaluate the performance of different model structures under conditions that are germane to modelling easement occurrence. Finally, we use a case study of conservation easements in Montana and Idaho to illustrate differences in inference that arise from different modelling choices.

Single-season Occupancy models

Typical occupancy studies are based on a design wherein a number of survey units are visited by trained observers multiple times during the course of a study period and the presence or absence of the target species is recorded. Observed absences may be the result of the species being truly absent from the location or the species being present, but not detected (MacKenzie et al., 2002). As such, data from occupancy surveys arise from the interaction of the ecological process that determines where a species occurs and an observation process that determines whether the observer records the species when it is present. Multiple visits to a site help distinguish between the two (Johnson, Conn, Hooten, Ray, & Pond, 2013). In situations where multiple visits to the same site is not possible (e.g., spatially extensive monitoring programs, remote sampling locations where repeat visits are cost prohibitive), spatial sub-units of the primary location have been used as a means of trading space for time with each spatial sub-unit treates as a repeat visit of the primary sample location (Kendall & White, 2009; pavlacky2012; Crosby & Porter, 2018).

We consider the issue of non-reporting in spatial databases of conservation action to be analogous to this latter situation. Consider a US Census tract which does not, according to the NCED, contain an easement. This may be because social-ecological conditions within the tract preclude an easement or because easement holders have chosen not to report them. There are a variety of reasons that voluntary conservation actions may be inconsistently reported. Conservation organizations, typical easement holders, may withhold information on easement location due to privacy concerns of the partnering landowner, legal restrictions, or fear of unintended use of the information (Rissman et al. 2017). Privacy concerns, the potential for trespassing, fear of additional monitoring or compliance burdens also prevent the easement grantor (i.e. landowner) may curtail the sharing of information by private landowners (Rissman et al. 2017).

We adapt the single-season occupancy model of (MacKenzie et al., 2002) to the nested geographies of the US Census by letting the true (but partially observed) occupancy status of $tract = 1, \ldots, n$ be z_t, where $z_t = 1$ if the tract has an easement and 0 if not. This is the true occupancy state, but is only partially observable. Within each tract there are $block\ groups = 1, \ldots, j$ that serve as nested subunits analagous to those used in broad-scale wildlife monitoring efforts. Assuming that individuals do not report non-existent easements (i.e., no false positives), the observed occupancy, $y_{t,bg}$ of $blockgroup_j$ within $tract_i$ will be 0 if $z_{tract} = 0$, but may be 1 (reported) or 0 (not-reported) if $z_{tract} = 1$. This gives rise to the standard specification for the single season occupncy model (e.g., Royle & Dorazio, 2008):

$$[y_{ij}|z_i, p_{ij}] = \text{Bernoulli}(z_i, p_{ij})$$
$$[z_i|\psi_i] = \text{Bernoulli}(\psi_i)$$

where p_{ij} is the probability of reporting an easement (given that it occurs) for block group j in tract i and ψ_i is the probability that an easement occurs in tract i. These probabilities can then be used to estimate the likelihood that a particular pattern of detections/nondetections occur. For example, consider tract i with a detection history of 101 indicating that two of the three block groups contain an easment. Based on that detection history, the likelihood for that tract would be:

$$\psi_i p_{i1}(1 - p_{i2})p_{i3}$$

indicating that tract i was occupied with probability ψ_i and either reported (p_{i1} and p_{i3}) or not ($1 - p_{i2}$). Relatedly, for tract k with a detection history of 000, the likelihood for this tract would be:

$$\psi_k \prod_{j=1}^{3}(1 - p_{kj}) + (1 - \psi_k)$$

indicating that tract k had an easement, but that easement went unreported at each block group (with probability $\psi_k \prod_{j=1}^{3}(1 - p_{kj})$) or did not have an easement (with probability $1 - \psi_k$).

These likelihoods can be used to evaluate hypotheses regarding the key social, institutional, or ecological predictors associated with likelihood of a location having an easement (ψ) while attempting to account for variables that may impact the likelihood that the easement is reported (p) by introducing covariates to explain variaion in ψ and p such that:

$$l(\psi_i) = \mathbf{x}_i \boldsymbol{\gamma} + \epsilon_\psi$$
$$l(p_{ij}) = \mathbf{x}_{ij}^{'} \boldsymbol{\beta} + \epsilon_p$$

where $l(\cdot)$ is a link function (typically logit) to constrain ψ_i and p_{ij} to the [0,1] interval and $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are vectors of regression coefficients relating covariates $\mathbf{x}_i$ and $\mathbf{x}_{ij}^{'}$ to ψ and p, respectively and ϵ are independent errors.

Substituting space for time

Because the datasets we envision here are generally a static snapshot of conservation behaviors, we are forced to use spatial subunits as a means of resampling a given tract. The closure assumption of the occupancy model presented here implies that if a sample unit (i.e., tract) is occupied then each subunit should have a non-negligible probability of detection (Hines et al., 2010; Pavlacky Jr, Blakesley, White, Hanni, & Lukacs, 2012). This is unlikely to be true for sessile organisms (Hines et al., 2010) and for conservation actions with spatialy fixed boundaries. In these instances, the observation process is composed of both the probability of detection (p_d) and the probability that an object of interest is available for sampling at the spatial subunit (p_a) (Guillera-Arroita, 2011). Revisiting tract i with a detection history of 101, the introduction of p_a results in the likelihood:

$$\psi_i p_{a_{i1}} p_{d_{i1}} \left[(1 - p_{a_{i2}}) + p_{a_{i2}}(1 - p_{d_{i2}})\right] p_{a_{i3}} p_{d_{i3}}$$

where the absence is now expressed as the probability that the block group is available, but the easement was unreported ($p_{a_{i2}}(1 - p_{d_{i2}})$) or the block group was unavailable ($1 - p_{a_{i2}}$). Unless p_d and p_a are modeled explicitly or surveys are designed explicitly (e.g., sampling with replacement) to control for variation in p_a, p_a and p_d may be confounded resulting in a situation where estimates of detection probability are biased low and occupancy are biased high (Kendall & White, 2009). This may be particularly problematic for efforts to characterize conservation action as not all locations within a tract may actually be available for the action. For example, conservation easements may not be possible in areas with no undeveloped land. Studies of wildlife occupancy have relied on the use of multiple detection methods or the ability to split transects into temporal "subsamples" based on the order in which locations are visited (e.g., Nichols et al., 2008; Hines et al., 2010; Pavlacky Jr et al., 2012); however, such approaches are rarely available when considering national databases of conservation actions.

Autocorrelation in occupancy or detection

Occupancy models assume that the occupancy state of one sample location is independent of the state of another sample location. Yet, in many instances sample locations near each other are more likely to be similar than those far apart (Tobler!!). Residual (i.e., after accounting for the effects of covariates) spatial autocorrelation may lead to biased estimates or overestimated precision leading to a number of efforts to account for spatial structure in the occupancy process (Johnson et al., 2013). The independence assumption can be relaxed by explicitly modeling the covariance matrix of the residuals, $\boldsymbol{\Sigma}$, as a function of the locations where the response variable was collected (Ver Hoef, Peterson, Hooten, Hanks, & Fortin, 2018). Autoregressive models provide a means of modeling $\mathbf{\Sigma}$ for areal data. Modeling the effect of covariates on occupancy then becomes:

$$l(\psi_i) = \mathbf{x}_i\boldsymbol{\gamma} + \mathbf{z} + \epsilon$$

where $\mathbf{z}$ is a latent spatial random error distributed as $\mathbf{z} \sim \mathcal{N}(0, \mathbf{\Sigma})$ and

$$\mathbf{\Sigma} = \tau(D - \rho W)^{-1}$$

resulting in the conditional autoregressive (CAR) model with a spatially varying precision paramater (τ), a diagonal matrix with the number of neighbors for a given location (D), a neighbor matrix (W), and the strength of spatial dependence (ρ). In addition to allowing relaxation of the independence assumption, autoregressive models may also improve estimation of parameters of interest by using estimates from nearby locations to smooth over potential measurement error such as that arrising from the detection process.

Methods

Our interest was in evaluating the bias that arises from (potentially unmodeled) variation in the probability of reporting an easement (given that it is there), the probability that a spatial location is available for an easement, and spatial autocorrelation in both occupancy and reporting probability.

The models

The model used to generate the data was:

$$\begin{aligned}
z_i &\sim \text{Bern}(\psi_i) \\
logit(\psi_i) &= \beta_0 + \beta\mathbf{X} + \phi_i \\
v_i|z_i &\sim \text{Bin}(n_{bg}, \alpha * z_i) \\
y_{i,j}|v_i z_i &\sim \text{Bern}(p_{i,j} * z_i * v_i) \\
logit(p_{i,j}) &= \gamma_0 + \gamma\mathbf{X} + \phi_{i,j}
\end{aligned}$$

where v=1 if available and 0 if not and n_bg is the number of block groups within a tract.

The models that I actually fit were as follows: Note that I did not explicitly model α because this would be difficult to do in-practice and my interest was in whether the approach would be robust to leaving that component out of the model (despite its role in the generating process)

Occupancy with CAR on both components

$$\begin{aligned}
z_i &\sim \text{Bern}(\psi_i) \\
logit(\psi_i) &= \beta_0 + \beta\mathbf{X} + \phi_i \\
y_{i,j}|z_i &\sim \text{Bern}(p_{i,j} * z_i) \\
logit(p_{i,j}) &= \gamma_0 + \gamma\mathbf{X} + \phi_{i,j}
\end{aligned}$$

Occupancy with CAR on occupancy only

$$\begin{aligned}
z_i &\sim \text{Bern}(\psi_i) \\
logit(\psi_i) &= \beta_0 + \beta\mathbf{X} + \phi_i \\
y_{i,j}|z_i &\sim \text{Bern}(p_{i,j} * z_i) \\
logit(p_{i,j}) &= \gamma_0 + \gamma\mathbf{X}
\end{aligned}$$

Occupancy with CAR on detection only

$$
\begin{aligned}
z_i &\sim \text{Bern}(\psi_i) \\
logit(\psi_i) &= \beta_0 + \beta\mathbf{X} \\
y_{i,j}|z_i &\sim \text{Bern}(p_{i,j} * z_i) \\
logit(p_{i,j}) &= \gamma_0 + \gamma\mathbf{X} + \phi_{i,j}
\end{aligned}
$$

Occupancy with no CAR component

$$
\begin{aligned}
z_i &\sim \text{Bern}(\psi_i) \\
logit(\psi_i) &= \beta_0 + \beta\mathbf{X} \\
y_{i,j}|z_i &\sim \text{Bern}(p_{i,j} * z_i) \\
logit(p_{i,j}) &= \gamma_0 + \gamma\mathbf{X}
\end{aligned}
$$

Binomial fit to the tract (i) level

$$
\begin{aligned}
y_i &\sim \text{Bern}(p_i) \\
logit(p) &= \beta_0 + \beta\mathbf{X} + \phi_i
\end{aligned}
$$

Generating covariate observations

We used the tracts and block groups for Iowa as the basis for developing our simulated datasets. We generated random data values for each of three predictors at the tract-level and two predictors at the block group-level. Because predictors may also be spatially autocorrelated, we simulated these data from $x_i \sim \mathcal{N}_{\|}(\boldsymbol{\mu} = 0, \boldsymbol{\Sigma} = \phi)$ where $\phi \sim \mathcal{N}(0, [\tau(D - \rho W)]^{-1})$ and τ, the spatially varying precision paramater is 1; ρ, the strength of spatial dependence is 0.3, D is a diagonal matrix containing the number of neighbors for a given location, and W is an adjacency matrix. We defined adjacency by determining the minimum distance necessary to ensure that all locations had at least one neighbor and considering to locations to be adjacent if they were within that distance from each other.

Latin Hypercube Sampling

We used a Latin hypercube design to evaluate model performance across a range of plausible parameter values. For each simulation replicate, we drew 300 uniformly distributed samples across the multi-dimensional space described by the limits in Table 1.

Generating observations

We simulated true easement occurrence for each sample generating a probability of easement occurrence for each tract by multiplying ψ (the mean occupancy probability for a given Latin hypercube sample [i.e., the intercept]) and three randomly generated regression coefficients by the tract-level design matrix and adding ϕ_Ψ where ϕ is described above and τ and ρ_Ψ were taken from the appropriate Latin hypercube sample. This resulted in true easement occurence (z_i) for each tract which we then converted to an observed occupancy dataset by generating an estimate of p (the block group reporting probability) by multiplying p (the mean reporting probability for a given Latin hypercube sample [i.e., the intercept]) and two randomly generated regression coefficents by the block group-level design matrix and adding ϕ_p with τ and

Table 1: Range of values used in Latin Hypercube sampler where Ψ is the probability of easement occurence, p is detection probability, $\rho_{...}$ is the strength of spatial autocorrelation for Ψ or p, τ is the precision for the conditional autoregressive term, and α is the probability that a location is available for an easement.

Variables	Lower	Upper
Ψ	0.20	0.80
p	0.30	0.98
ρ_Ψ	0.50	1.00
ρ_p	0.50	1.00
τ	0.10	1.00
α	0.20	0.80

ρ_p as above. Finally, because the probability of an easement being reported (conditional on its existence) depends on both the probability of reporting (p) and the probability that a block group is available for an easement (α), we simulated easement observations following $y_{i,j} \sim \text{Bern}(\alpha p)$.

Fitting models

We generated 50 replicates [only showing results from 4 here] of each Latin hypercube sample (1500 simulated datasets total) and fit each of five models (a standard occupancy model, a binomial model with a conditional autoregressive [CAR] term for spatial autocorrelation, an occupancy model with a CAR term for occurrence only, an occupancy model with a CAR term for reporting only, and an occupancy model with a CAR terms for both components) to the simulated dataset. All models were fit in R using rstan, a wrapper to the Stan language (stan models are at end of document). Models were fit using an adaptation parameter of 0.98 and and a maximum treedepth of 16 with four chains each run for 3700 iterations (with a warmup period of 3200 iterations). We calcaulated the relative bias for the intercepts and regression coefficients as:

$$RelBias = \frac{\hat{\beta} - \beta}{|\beta|}$$

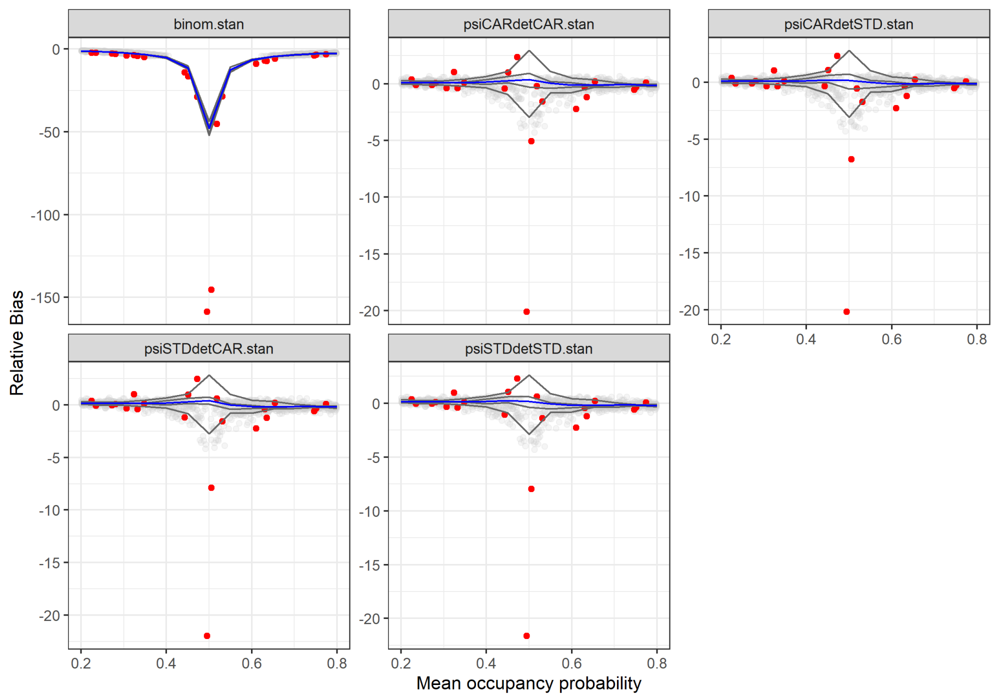

Figure 1: Relative bias of the posterior estimates of the intercept in the binomial model and the intercept in the occupancy component of the occupancy modes for each simulated dataset. Gray dots are median estimates of relative bias from each model run, gray lines are replicate-specific the relationship between median values of relative bias and the simulated mean occupancy probability, blue lines depict the relationship between median values of relative bias and occupancy probability across all simulation replicates, red dots indicate simulation runs where both the probability of detection and the probability of availibility were in the lower decile.

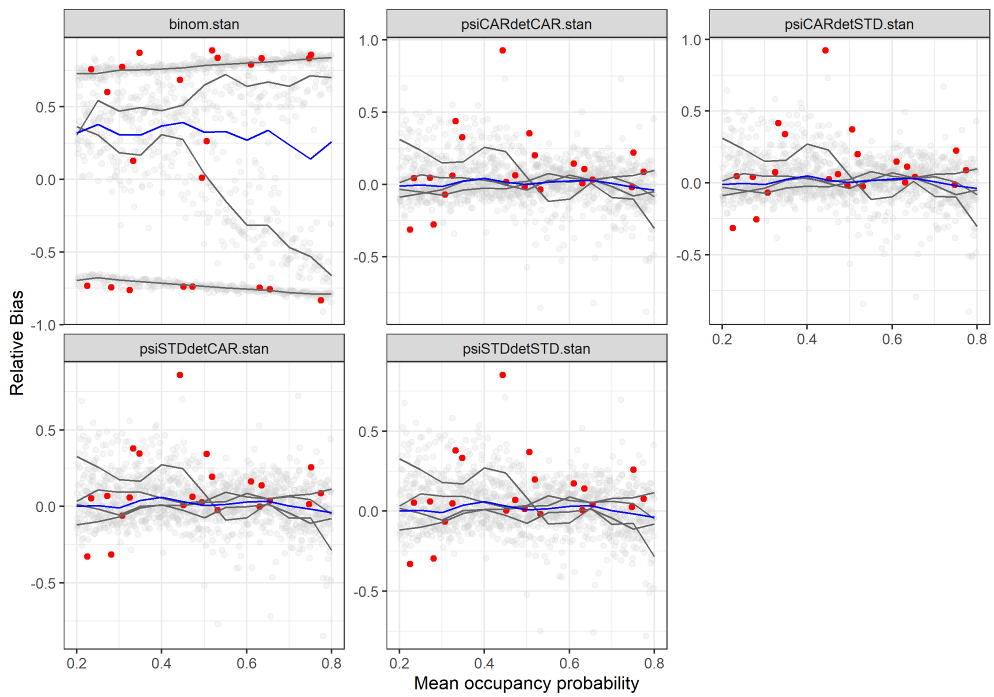

Figure 2: Relative bias of the posterior estimates of the regression coefficients in the binomial model and the intercept in the occupancy component of the occupancy modes for each simulated dataset. Gray dots are median estimates of relative bias from each model run, gray lines are replicate-specific the relationship between median values of relative bias and the simulated mean occupancy probability, blue lines depict the relationship between median values of relative bias and occupancy probability across all simulation replicates, red dots indicate simulation runs where both the probability of detection and the probability of availibility were in the lower decile.

Stan models

Binomial

```
functions{
/**
* Return log probability of a unit-scale proper conditional autoregressive
  * (CAR) prior with a sparse representation for the adjacency matrix
  *
  * @param phi Vector containing the parameters with a CAR prior
  * @param alpha Dependence (usually spatial) parameter for the CAR prior (real)
  * @param W_sparse Sparse representation of adjacency matrix (int array)
  * @param n Length of phi (int)
  * @param W_n Number of adjacent pairs (int)
  * @param D_sparse Number of neighbors for each location (vector)
  * @param lambda Eigenvalues of D^{-1/2}*W*D^{-1/2} (vector)
  *
  * @return Log probability density of CAR prior up to additive constant
  */
  real sparse_car_lpdf(vector phi, real alpha,
    int[,] W_sparse, vector D_sparse, vector lambda, int nobs, int W_n) {
      row_vector[nobs] phit_D; // phi' * D
      row_vector[nobs] phit_W; // phi' * W
      vector[nobs] ldet_terms;

      phit_D = (phi .* D_sparse)';
      phit_W = rep_row_vector(0, nobs);
      for (i in 1:W_n) {
        phit_W[W_sparse[i, 1]] += phi[W_sparse[i, 2]];
        phit_W[W_sparse[i, 2]] += phi[W_sparse[i, 1]];
      }

      for (i in 1:nobs) ldet_terms[i] = log1m(alpha * lambda[i]);
      return 0.5 * (sum(ldet_terms)
                    - (phit_D * phi - alpha * (phit_W * phi)));
  }
}
data {
  // site-level occupancy covariates
  int<lower = 1> n_site;
  int<lower = 1> m_psi;
  matrix[n_site, m_psi] X_tct;


  // summary of whether species is known to be present at each site
  int<lower = 0, upper = 1> any_seen[n_site];

  // number of surveys at each site
  int<lower = 0> n_survey[n_site];

  matrix<lower = 0, upper = 1>[n_site, n_site] W_tct; //adjacency matrix tract
  int W_n_tct; //number of adjacent pairs
}
transformed data {
```

```
  int W_sparse_occ[W_n_tct, 2];   // adjacency pairs
  vector[n_site] D_sparse_occ;     // diagonal of D (number of neigbors for each site)

  vector[n_site] lambda_occ;       // eigenvalues of invsqrtD * W * invsqrtD
  { // generate sparse representation for W
  int counter_occ;
  counter_occ = 1;
  // loop over upper triangular part of W to identify neighbor pairs
    for (i in 1:(n_site - 1)) {
      for (j in (i + 1):n_site) {
        if (W_tct[i, j] == 1) {
          W_sparse_occ[counter_occ, 1] = i;
          W_sparse_occ[counter_occ, 2] = j;
          counter_occ += 1;
        }
      }
    }
  }
  for (i in 1:n_site) D_sparse_occ[i] = sum(W_tct[i]);
  {
    vector[n_site] invsqrtD_occ;
    for (i in 1:n_site) {
      invsqrtD_occ[i] = 1 / sqrt(D_sparse_occ[i]);
    }
    lambda_occ = eigenvalues_sym(quad_form(W_tct, diag_matrix(invsqrtD_occ)));
  }

}
parameters {
  vector[n_site] phi_occ;
  real<lower = 0, upper =0.999> alpha_occ;
  real<lower = 0> sigma_occ;
  vector[m_psi] beta_psi;
}
transformed parameters {
  vector[n_site] logit_psi = X_tct * beta_psi + phi_occ * sigma_occ;
}
model {
   phi_occ ~ sparse_car(alpha_occ, W_sparse_occ, D_sparse_occ, lambda_occ, n_site, W_n_tct);
  alpha_occ ~ beta(4,1);
  sigma_occ ~ normal(0, 1.5);
  beta_psi ~ student_t(7.763,0, 1.566);

  any_seen ~ binomial_logit(n_survey, logit_psi);
}
```

Occupancy without CAR

Note that although this code has the functions to estimate ϕ, the actual model does not do so.

```
functions{
/**
* Return log probability of a unit-scale proper conditional autoregressive
  * (CAR) prior with a sparse representation for the adjacency matrix
  *
  * @param phi Vector containing the parameters with a CAR prior
  * @param alpha Dependence (usually spatial) parameter for the CAR prior (real)
  * @param W_sparse Sparse representation of adjacency matrix (int array)
  * @param n Length of phi (int)
  * @param W_n Number of adjacent pairs (int)
  * @param D_sparse Number of neighbors for each location (vector)
  * @param lambda Eigenvalues of D^{-1/2}*W*D^{-1/2} (vector)
  *
  * @return Log probability density of CAR prior up to additive constant
  */
  real sparse_car_lpdf(vector phi, real alpha,
    int[,] W_sparse, vector D_sparse, vector lambda, int nobs, int W_n) {
      row_vector[nobs] phit_D; // phi' * D
      row_vector[nobs] phit_W; // phi' * W
      vector[nobs] ldet_terms;

      phit_D = (phi .* D_sparse)';
      phit_W = rep_row_vector(0, nobs);
      for (i in 1:W_n) {
        phit_W[W_sparse[i, 1]] += phi[W_sparse[i, 2]];
        phit_W[W_sparse[i, 2]] += phi[W_sparse[i, 1]];
      }

      for (i in 1:nobs) ldet_terms[i] = log1m(alpha * lambda[i]);
      return 0.5 * (sum(ldet_terms)
                    - (phit_D * phi - alpha * (phit_W * phi)));
  }
}
data {
  // site-level occupancy covariates
  int<lower = 1> n_site;
  int<lower = 1> m_psi;
  matrix[n_site, m_psi] X_tct;

  // survey-level detection covariates
  int<lower = 1> total_surveys;
  int<lower = 1> m_p;
  matrix[total_surveys, m_p] X_bg;

  // survey level information
  int<lower = 1, upper = n_site> site[total_surveys];
  int<lower = 0, upper = 1> y[total_surveys];
  int<lower = 0, upper = total_surveys> start_idx[n_site];
  int<lower = 0, upper = total_surveys> end_idx[n_site];

  // summary of whether species is known to be present at each site
```

```
  int<lower = 0, upper = 1> any_seen[n_site];

  // number of surveys at each site
  int<lower = 0> n_survey[n_site];

}
parameters {
  vector[m_psi] beta_psi;
  vector[m_p] beta_p;
}
transformed parameters {
  vector[total_surveys] logit_p = X_bg * beta_p;
  vector[n_site] logit_psi = X_tct * beta_psi;
}
model {
  vector[n_site] log_psi = log_inv_logit(logit_psi);
  vector[n_site] log1m_psi = log1m_inv_logit(logit_psi);

  beta_psi ~ student_t(7.763,0, 1.566);
  beta_p ~ student_t(7.763,0, 1.566);

  for (i in 1:n_site) {
    if (n_survey[i] > 0) {
      if (any_seen[i]) {
        // site is occupied
        target += log_psi[i]
                  + bernoulli_logit_lpmf(y[start_idx[i]:end_idx[i]] |
                                         logit_p[start_idx[i]:end_idx[i]]);
      } else {
        // site may or may not be occupied
        target += log_sum_exp(
          log_psi[i] + bernoulli_logit_lpmf(y[start_idx[i]:end_idx[i]] |
                                            logit_p[start_idx[i]:end_idx[i]]),
          log1m_psi[i]
        );
      }
    }
  }
}
```

Occupancy with CAR on both components

```
functions{
/**
* Return log probability of a unit-scale proper conditional autoregressive
  * (CAR) prior with a sparse representation for the adjacency matrix
  *
  * @param phi Vector containing the parameters with a CAR prior
  * @param alpha Dependence (usually spatial) parameter for the CAR prior (real)
  * @param W_sparse Sparse representation of adjacency matrix (int array)
  * @param n Length of phi (int)
  * @param W_n Number of adjacent pairs (int)
  * @param D_sparse Number of neighbors for each location (vector)
  * @param lambda Eigenvalues of D^{-1/2}*W*D^{-1/2} (vector)
  *
  * @return Log probability density of CAR prior up to additive constant
  */
  real sparse_car_lpdf(vector phi, real alpha,
    int[,] W_sparse, vector D_sparse, vector lambda, int nobs, int W_n) {
      row_vector[nobs] phit_D; // phi' * D
      row_vector[nobs] phit_W; // phi' * W
      vector[nobs] ldet_terms;

      phit_D = (phi .* D_sparse)';
      phit_W = rep_row_vector(0, nobs);
      for (i in 1:W_n) {
        phit_W[W_sparse[i, 1]] += phi[W_sparse[i, 2]];
        phit_W[W_sparse[i, 2]] += phi[W_sparse[i, 1]];
      }

      for (i in 1:nobs) ldet_terms[i] = log1m(alpha * lambda[i]);
      return 0.5 * (sum(ldet_terms)
                    - (phit_D * phi - alpha * (phit_W * phi)));
  }
}
data {
  // site-level occupancy covariates
  int<lower = 1> n_site;
  int<lower = 1> m_psi;
  matrix[n_site, m_psi] X_tct;

  // survey-level detection covariates
  int<lower = 1> total_surveys;
  int<lower = 1> m_p;
  matrix[total_surveys, m_p] X_bg;

  // survey level information
  int<lower = 1, upper = n_site> site[total_surveys];
  int<lower = 0, upper = 1> y[total_surveys];
  int<lower = 0, upper = total_surveys> start_idx[n_site];
  int<lower = 0, upper = total_surveys> end_idx[n_site];

  // summary of whether species is known to be present at each site
  int<lower = 0, upper = 1> any_seen[n_site];
```

```
  // number of surveys at each site
  int<lower = 0> n_survey[n_site];

  //adjacency data
  matrix<lower = 0, upper = 1>[n_site, n_site] W_tct; //adjacency matrix tract
  int W_n_tct; //number of adjacent pairs
  matrix<lower = 0, upper = 1>[total_surveys, total_surveys] W_bg; //adjacency matrix bg
  int W_n_bg; //number of adjacent pairs bg
  //real<lower = 0, upper =1> alpha_occ;
  //real<lower = 0, upper =1> alpha_det;

}
transformed data {
  int W_sparse_occ[W_n_tct, 2];   // adjacency pairs
  int W_sparse_det[W_n_bg, 2];
  vector[n_site] D_sparse_occ;     // diagonal of D (number of neigbors for each site)
  vector[total_surveys] D_sparse_det;

  vector[n_site] lambda_occ;       // eigenvalues of invsqrtD * W * invsqrtD
  vector[total_surveys] lambda_det;
  { // generate sparse representation for W
  int counter_occ;
  counter_occ = 1;
  // loop over upper triangular part of W to identify neighbor pairs
    for (i in 1:(n_site - 1)) {
      for (j in (i + 1):n_site) {
        if (W_tct[i, j] == 1) {
          W_sparse_occ[counter_occ, 1] = i;
          W_sparse_occ[counter_occ, 2] = j;
          counter_occ += 1;
        }
      }
    }
  }
  for (i in 1:n_site) D_sparse_occ[i] = sum(W_tct[i]);
  {
    vector[n_site] invsqrtD_occ;
    for (i in 1:n_site) {
      invsqrtD_occ[i] = 1 / sqrt(D_sparse_occ[i]);
    }
    lambda_occ = eigenvalues_sym(quad_form(W_tct, diag_matrix(invsqrtD_occ)));
  }
  { // generate sparse representation for W_det
  int counter_det;
  counter_det = 1;
  // loop over upper triangular part of W to identify neighbor pairs
    for (i in 1:(total_surveys - 1)) {
      for (j in (i + 1):total_surveys) {
        if (W_bg[i, j] == 1) {
          W_sparse_det[counter_det, 1] = i;
          W_sparse_det[counter_det, 2] = j;
          counter_det += 1;
        }
```

```
      }
    }
  }
  for (i in 1:total_surveys) D_sparse_det[i] = sum(W_bg[i]);
  {
    vector[total_surveys] invsqrtD_det;
    for (i in 1:total_surveys) {
      invsqrtD_det[i] = 1 / sqrt(D_sparse_det[i]);
    }
    lambda_det = eigenvalues_sym(quad_form(W_bg, diag_matrix(invsqrtD_det)));
  }
}

parameters {
  vector[n_site] phi_occ;
  vector[total_surveys] phi_det;
  real<lower = 0, upper =0.999> alpha_occ;
  real<lower = 0, upper =0.999> alpha_det;
  real<lower = 0> sigma_occ;
  real<lower = 0> sigma_det;

  vector[m_psi] beta_psi;
  vector[m_p] beta_p;
}
transformed parameters {
  vector[total_surveys] logit_p = X_bg * beta_p + phi_det * sigma_det;
  vector[n_site] logit_psi = X_tct * beta_psi + phi_occ * sigma_occ;
}
model {
  vector[n_site] log_psi = log_inv_logit(logit_psi);
  vector[n_site] log1m_psi = log1m_inv_logit(logit_psi);

  phi_occ ~ sparse_car(alpha_occ, W_sparse_occ, D_sparse_occ, lambda_occ, n_site, W_n_tct);
  alpha_occ ~ beta(4,1);
  sigma_occ ~ normal(0, 1.5);
  phi_det ~ sparse_car(alpha_det, W_sparse_det, D_sparse_det, lambda_det, total_surveys, W_n_bg);
  alpha_det ~ beta(4,1);
  sigma_det ~ normal(0, 1.5);

  beta_psi ~ student_t(7.763,0, 1.566);
  beta_p ~ student_t(7.763,0, 1.566);

  for (i in 1:n_site) {
    if (n_survey[i] > 0) {
      if (any_seen[i]) {
        // site is occupied
        target += log_psi[i]
                  + bernoulli_logit_lpmf(y[start_idx[i]:end_idx[i]] |
                                         logit_p[start_idx[i]:end_idx[i]]);
      } else {
        // site may or may not be occupied
        target += log_sum_exp(
          log_psi[i] + bernoulli_logit_lpmf(y[start_idx[i]:end_idx[i]] |
                                            logit_p[start_idx[i]:end_idx[i]]),
```

```
          log1m_psi[i]
        );
      }
    }
  }
}
```


Bailey, L. L., MacKenzie, D. I., & Nichols, J. D. (2014). Advances and applications of occupancy models. *Methods in Ecology and Evolution*, *5*(12), 1269–1279.

Cheever, F., & MaLaughlin, N. A. (2014). An introduction to conservation easements in the united states: A simple concept and a complicated mosaic of law. *JL Prop. & Soc'y*, *1*, 107.

Crosby, A. D., & Porter, W. F. (2018). A spatially explicit, multi-scale occupancy model for large-scale population monitoring. *The Journal of Wildlife Management*, *82*(6), 1300–1310. doi:10.1002/jwmg.21466

Gu, W., & Swihart, R. K. (2004). Absent or undetected? Effects of non-detection of species occurrence on wildlife–habitat models. *Biological Conservation*, *116*(2), 195–203.

Guillera-Arroita, G. (2011). Impact of sampling with replacement in occupancy studies with spatial replication. *Methods in Ecology and Evolution*, *2*(4), 401–406. doi:10.1111/j.2041-210X.2011.00089.x

Guillera-Arroita, J. J. A. M., Gurutzeta AND Lahoz-Monfort. (2014). Ignoring imperfect detection in biological surveys is dangerous: A response to 'fitting and interpreting occupancy models'. *PLOS ONE*, *9*(7), 1–14. doi:10.1371/journal.pone.0099571

Hines, J. E., Nichols, J. D., Royle, J. A., MacKenzie, D. I., Gopalaswamy, A. M., Kumar, N. S., & Karanth, K. U. (2010). Tigers on trails: Occupancy modeling for cluster sampling. *Ecological Applications*, *20*(5), 1456–1466. doi:10.1890/09-0321.1

Johnson, D. S., Conn, P. B., Hooten, M. B., Ray, J. C., & Pond, B. A. (2013). Spatial occupancy models for large data sets. *Ecology*, *94*(4), 801–808.

Kellner, R. K., Kenneth F. AND Swihart. (2014). Accounting for imperfect detection in ecology: A quantitative review. *PLOS ONE*, *9*(10), 1–8. doi:10.1371/journal.pone.0111436

Kendall, W. L., & White, G. C. (2009). A cautionary note on substituting spatial subunits for repeated temporal sampling in studies of site occupancy. *Journal of Applied Ecology*, *46*(6), 1182–1188.

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Andrew Royle, J., & Langtimm, C. A. (2002). ESTIMATING site occupancy rates when detection probabilities are less than one. *Ecology*, *83*(8), 2248–2255. doi:10.1890/0012-9658(2002)083[2248:ESORWD]2.0.CO;2

MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L., & Hines, J. E. (2017). *Occupancy estimation and modeling: Inferring patterns and dynamics of species occurrence*. Elsevier.

Margules, C. R., & Pressey, R. L. (2000). Systematic conservation planning. *Nature*, *405*(6783), 243–253.

Nichols, J. D., Bailey, L. L., O'Connell Jr., A. F., Talancy, N. W., Campbell Grant, E. H., Gilbert, A. T., … Hines, J. E. (2008). Multi-scale occupancy estimation and modelling using multiple detection methods. *Journal of Applied Ecology*, *45*(5), 1321–1329. doi:10.1111/j.1365-2664.2008.01509.x

Pavlacky Jr, D. C., Blakesley, J. A., White, G. C., Hanni, D. J., & Lukacs, P. M. (2012). Hierarchical multi-scale occupancy estimation for monitoring wildlife populations. *The Journal of Wildlife Management*, *76*(1), 154–162.

Rissman, A. R., Owley, J., L'Roe, A., Morris, A., & Wardropper, C. (2017). Public access to spatial data on private-land conservation. *Ecology and Society*, *22*(2), 24.

Royle, J. A., & Dorazio, R. M. (2008). *Hierarchical modeling and inference in ecology: The analysis of data from populations, metapopulations and communities*. Elsevier.

Ver Hoef, J. M., Peterson, E. E., Hooten, M. B., Hanks, E. M., & Fortin, M.-J. (2018). Spatial autoregressive models for statistical inference from ecological data. *Ecological Monographs*, *88*(1), 36–59. doi:10.1002/ecm.1283

Welsh, D. B. A. D., Alan H. AND Lindenmayer. (2013). Fitting and interpreting occupancy models. *PLOS ONE*, *8*(1), 1–21. doi:10.1371/journal.pone.0052015

Williamson, M. A., Schwartz, M. W., & Lubell, M. N. (2018). Spatially explicit analytical models for social–ecological systems. *BioScience*, *68*(11), 885–895.