

Hierarchical models:

An application with multi-species
occupancy models

Bayesian computation in JAGS

Run through R with the library "R2jags"

Need to specify the following things

1. A text file with the model (must include the likelihood and the priors)
2. The data that will be used to estimate parameters in the model (data)
3. Initial values for the Markov chains (inits)
4. The parameters that we want to monitor (params)
5. The number of chains (nc)
6. The number of iterations for each chain to run (ni)
7. The number of iterations that should be discarded as a burn in (nb)
8. The number by which to thin the chains (nt)

Hierarchical models

Hierarchical modeling is a generalization of linear (and generalized linear) modeling in which regression coefficients are themselves given a model, whose parameters are also estimated from data

Gelman and Hill 2007

Hierarchical models

Hierarchical modeling is a generalization of linear (and generalized linear) modeling in which regression coefficients are themselves given a model, whose parameters are also estimated from data

$$y = mx + b + \varepsilon$$

$$\varepsilon \sim N(0, \sigma)$$

Hierarchical models

Hierarchical modeling is a generalization of linear (and generalized linear) modeling in which regression coefficients are themselves given a model, whose parameters are also estimated from data

$$y = mx + b_i + \varepsilon$$

$$\varepsilon \sim N(0, \sigma)$$

Hierarchical models

Hierarchical modeling is a generalization of linear (and generalized linear) modeling in which regression coefficients are themselves given a model, whose parameters are also estimated from data

$$y = mx + b_i + \varepsilon$$
$$\varepsilon \sim N(0, \sigma)$$
$$b_i \sim N(\mu_b, \sigma_b)$$

Hierarchical models

Hierarchical modeling is a generalization of linear (and generalized linear) modeling in which regression coefficients are themselves given a model, whose parameters are also estimated from data

Called the hyper-distribution

$$\longrightarrow b_i \sim N(\mu_b, \sigma_b)$$

Hierarchical models

- Hierarchical models are really just a random effects model!
- Technically, all random effects models are hierarchical models
- In practice we want to use hierarchical models to separate out important biological processes *and* sampling processes

Multispecies occupancy modeling

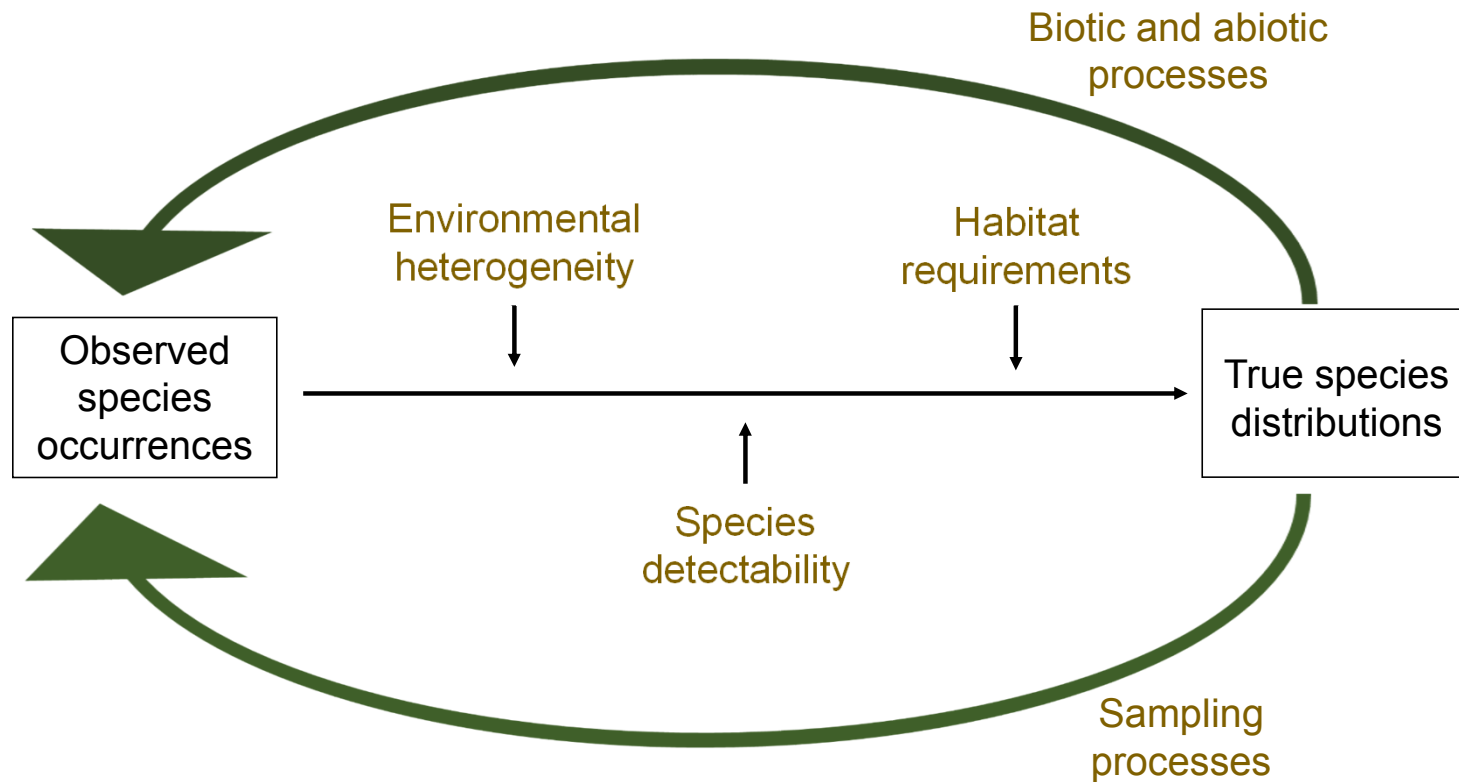
Why is it challenging to make inferences species distributions, abundances, and richness?

It is virtually impossible to count every individual/species during sampling

Hierarchical models can help!

- Separate out the processes affecting a species occurrence in a particular location from non-detection
- Account for rare and/or elusive species by treating species as random effects
 - Still accounts for the fact that species respond to their environments in different ways
 - But also borrows strength from the ensemble

Multi-species occurrence models



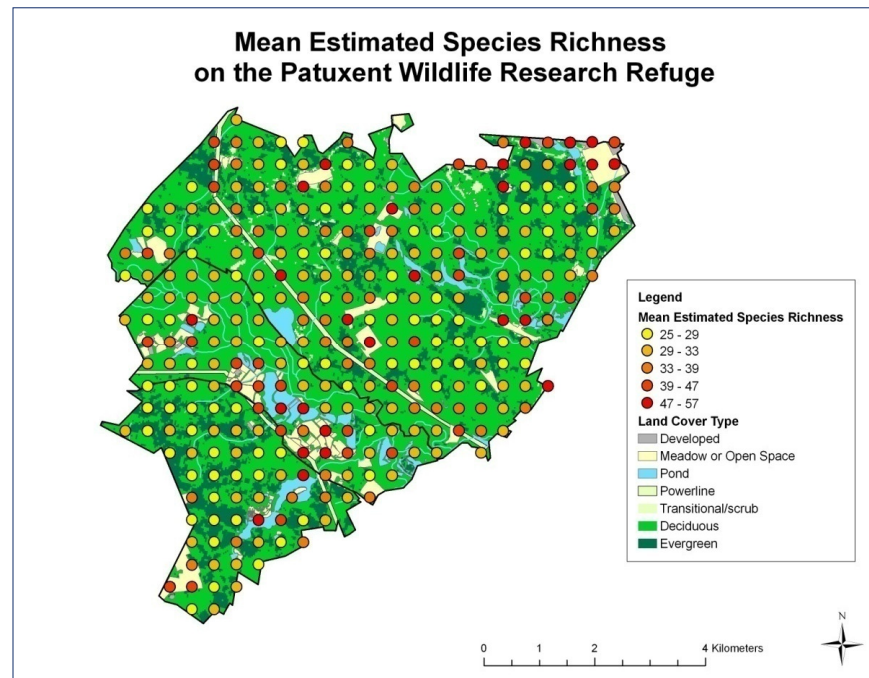
What is multi-species modeling?

Species	Location																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
A	x		x			x		x	x			x		x		x				
B			x		x	x				x		x		x			x			
C	x	x		x		x	x			x			x	x				x		x
D		x		x	x	x				x	x			x	x	x	x		x	
E	x	x	x		x	x										x		x		x
F			x	x							x			x			x	x		x
G		x		x	x								x					x		
H	x										x					x		x		
I								x					x							

- Many surveys produce data on multiple species
- Link single species occupancy models at the community level

Why multi-species modeling?

Models reflect how species are distributed and how they assemble to form communities



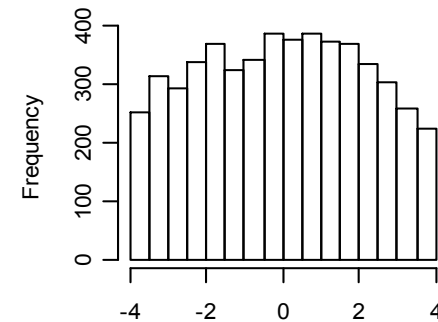
Why multi-species modeling?

Increased precision in estimates of occupancy and covariate effects, especially for rare species

Cerulean warbler



Single species model

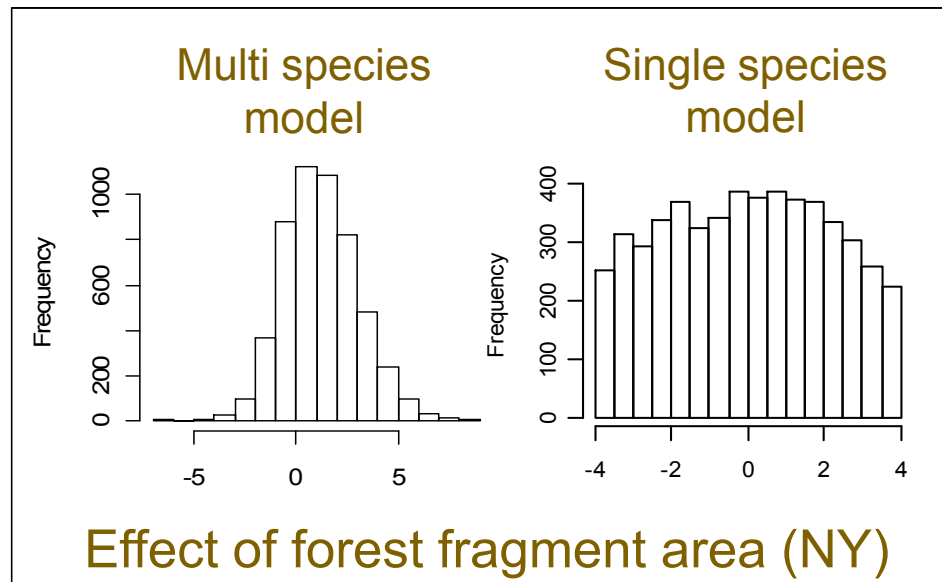


Effect of forest fragment area (NY)

Why multi-species modeling?

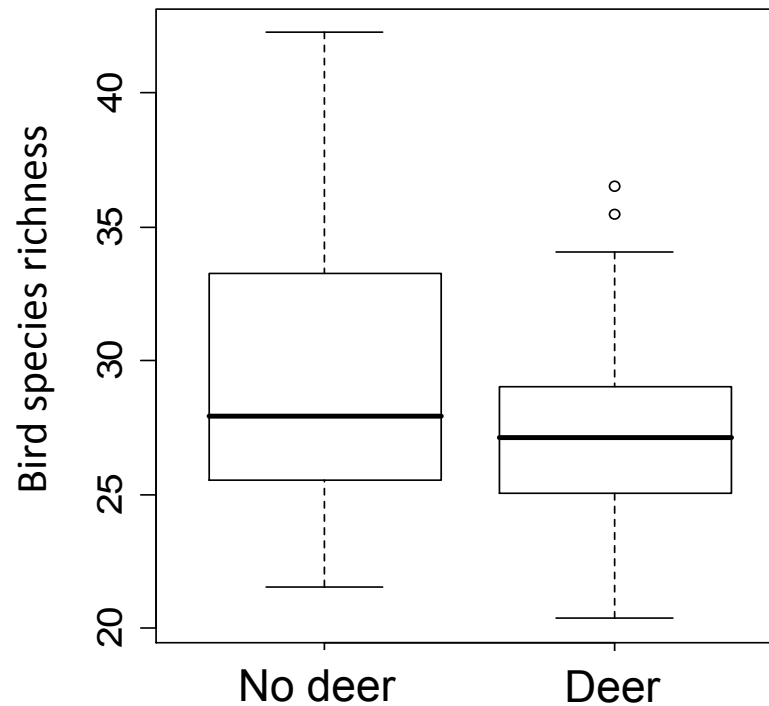
Increased precision in estimates of occupancy and covariate effects, especially for rare species

Cerulean warbler



Why multi-species modeling?

Produce estimates of the “community effect”



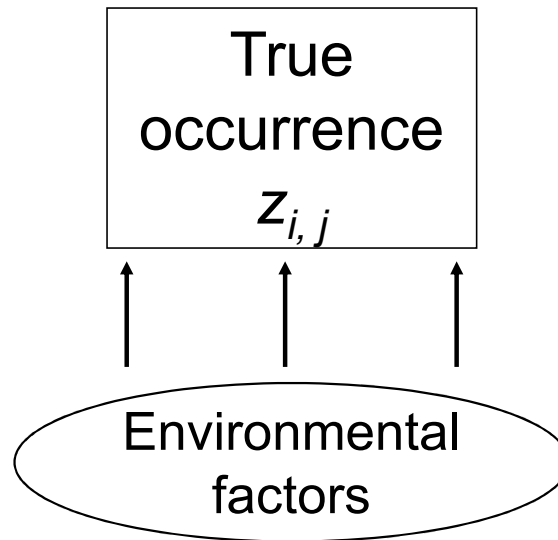
Model process

True
occurrence
 $z_{i,j}$

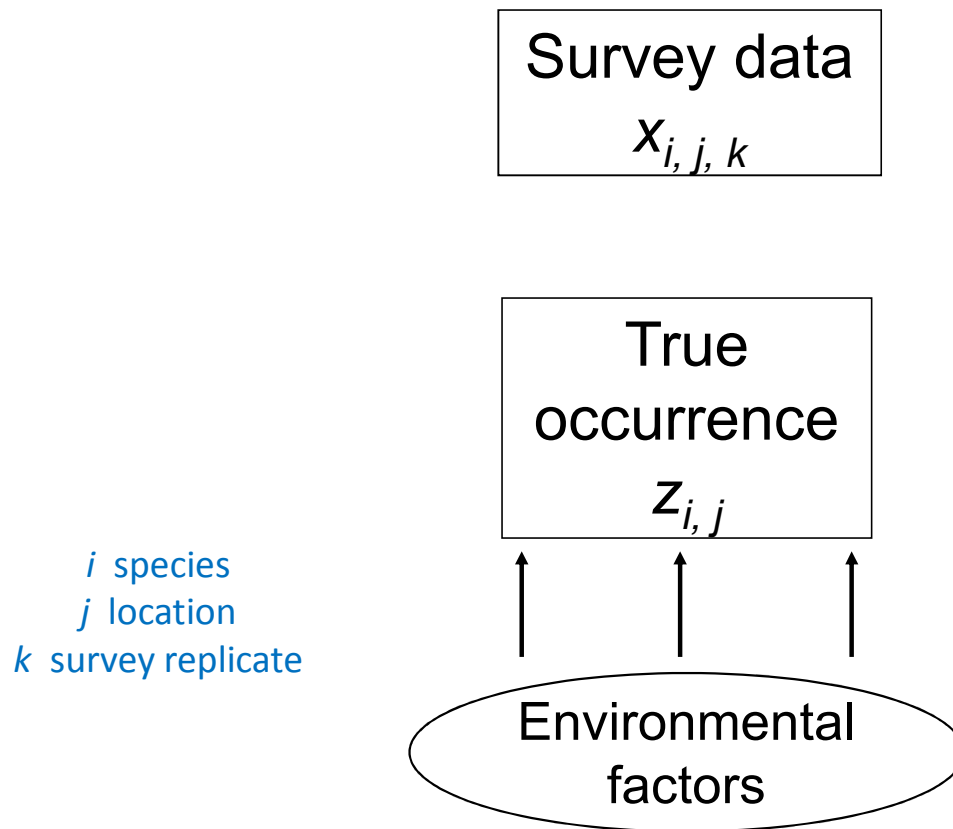
i species
 j location

Model process

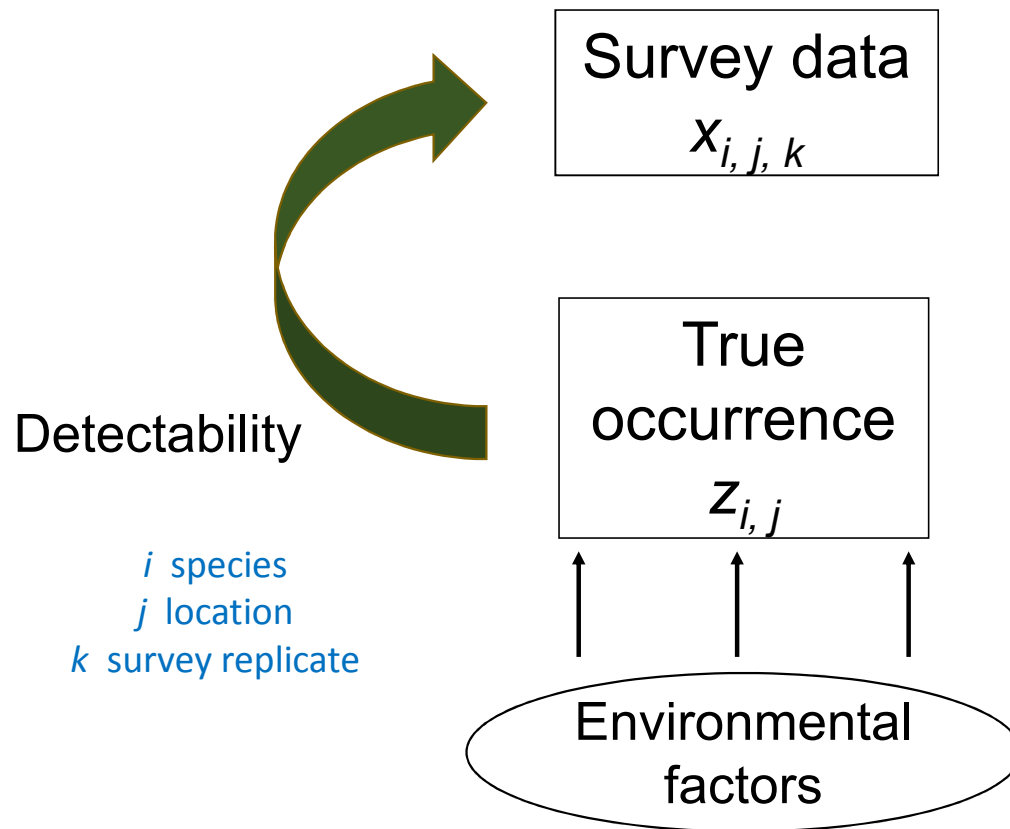
i species
j location



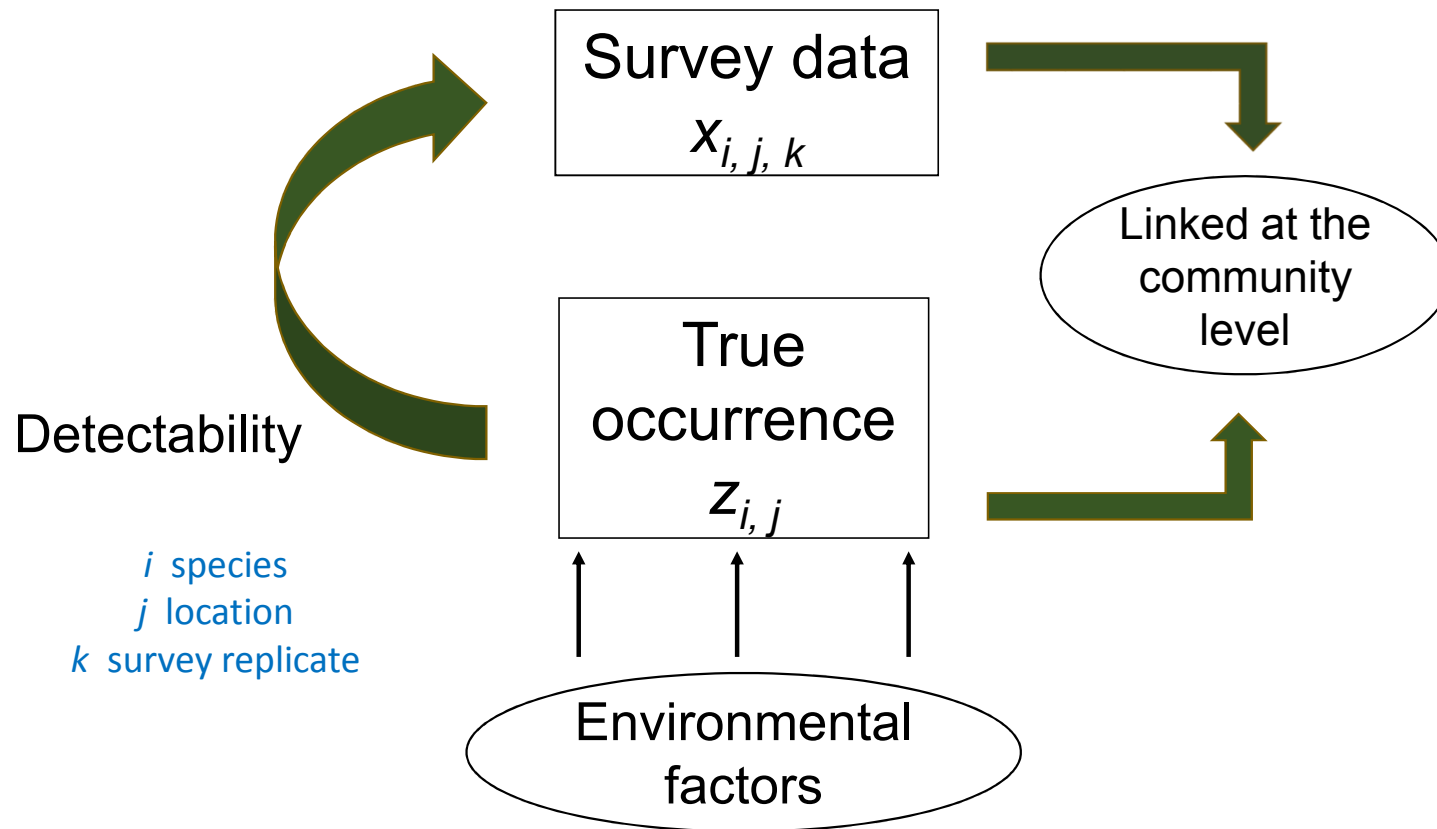
Model process



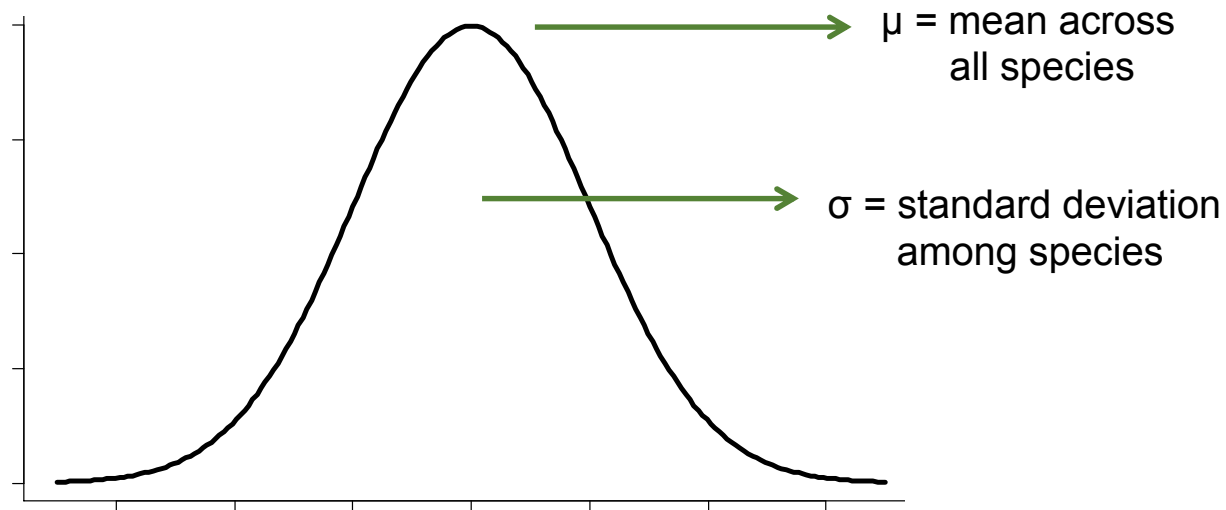
Model process



Model process



What does it mean to link species models?



Assume that species covariates are drawn from a common – community-level – distribution

What does it mean to link species models?

Effect of forest fragment area for species i (α_i)

$$\alpha_i \sim N(\mu_\alpha, \sigma_\alpha)$$

μ_α = mean effect of forest area on all species

σ_α = standard deviation of the effect of forest area among all species

Single species occupancy model

Process model

$$z_j \sim \text{Bernoulli}(\psi_j)$$

$$\text{logit}(\psi_j) = \alpha_0 + \alpha_1 \cdot x_j$$

z_j = True occurrence of the species at location j

($z_j=1$ if occupied)

ψ_i = Occurrence probability at location j

x_j = Covariate at j (e.g., habitat variable)

Single species occupancy model

Observation model

$$y_{j,k} \mid z_j \sim \text{Bernoulli}(z_j \cdot p_{j,k})$$

$$\text{logit}(p_{j,k}) = \beta_0 + \beta_1 \cdot w_{j,k}$$

$y_{j,k}$ = Detection/nodetection data at sampling location j during replicate k ($y_{j,k} = 1$ if detected)

$p_{j,k}$ = Detection probability at sampling location i during k

$w_{j,k}$ = Covariate affecting detection at j during replicate k (e.g., weather, time of day)

Assumptions

- **Single season** occupancy models – assume the sites are “closed” to changes in occupancy. Thus, a site is either considered occupied or not during the season and cannot change status
- **Occupancy probability** is assumed constant across sites or we can incorporate covariates to model differences
- **Detection probability** can be constant or modeled as a function of covariates as well
- **Detection** among sites is independent and there are *no false positives*

Multi-species occupancy model

Process model

$$z_{i,j} \sim \text{Bernoulli}(\psi_{i,j})$$

$$\text{logit}(\psi_{i,j}) = \alpha 0_i + \alpha 1_i \cdot x_j$$

$z_{i,j}$ = True occurrence of species i at location j

($z_{i,j}=1$ if species i occurs at j)

$\psi_{i,j}$ = Occurrence probability of species i at location j

x_j = Covariate at j (e.g., habitat variable)

Multi-species occupancy model

Observation model

$$y_{i,j,k} \mid z_i \sim \text{Bernoulli}(z_i \cdot p_{i,j,k})$$

$$\text{logit}(p_{i,j,k}) = \beta 0_i + \beta 1_i \cdot w_{j,k}$$

$y_{i,j,k}$ = Detection/nodetection data at sampling of species i at location j during replicate k ($y_{i,k} = 1$ if detected)

$p_{i,j,k}$ = Detection probability of species i at sampling location j during k

$w_{j,k}$ = Covariate affecting detection at j during replicate k (e.g., weather, time of day)

Multi-species occupancy model

- Link species by assuming that each of the covariates are drawn from a common – community-level – distribution

$$\alpha_{0i} \sim N(\mu_{\alpha_0}, \sigma_{\alpha_0})$$

where

μ_{α_0} = mean baseline occupancy across all species (logit scale)

σ_{α_0} = standard deviation in baseline occupancy across species

Multi-species occupancy model

Example

- Interest in modeling the effects of forest fragmentation in the Hudson River Valley (upstate NY) on the bird community.
- 72 randomly selected sites in fragmented forests were surveyed using point counts over a two year period.
- Each site was visited three times over a short time period when the community was considered “closed”.
- 78 species were observed. Of these, the data for 32 species were particularly sparse with less than 20 detections each over the entirety of the sampling season.

Multi-species occupancy model

Example

- Hypothesis: forest interior species would have highest occupancy in forest patches with large area and P/A ratios.
- Other species would be more mixed in preferences but overall, we expect larger patches with low P/A values to have the greatest richness

Multi-species occupancy model

Process model: bird community

$$z_{i,j} \sim \text{Bernoulli}(\psi_{i,j})$$

$$\text{logit}(\psi_{i,j}) = \alpha 0_i + \alpha 1_i \cdot \text{prem}_j + \alpha 2_i \cdot \text{area}_j \\ + \alpha 3_i \cdot P/A_j$$

Multi-species occupancy model

Observation model: bird community

$$y_{i,j,k} \mid z_i \sim \text{Bernoulli}(z_i \cdot p_{i,j,k})$$

$$\text{logit}(p_{i,j,k}) = \beta 0_i + \beta 1_i \cdot \text{date}_{j,k} + \beta 2_i \cdot \text{date}_{j,k}^2 \\ + \beta 3_i \cdot \text{year}_{j,k}$$

Multi-species occupancy model

- Assume each covariates $\{\alpha_{0i}, \alpha_{1i}, \alpha_{2i}, \alpha_{3i}, \beta_{0i}, \beta_{1i}, \beta_{2i}, \beta_{3i}\}$ are drawn from a common – community-level – distribution

$$\alpha_{1i} \sim N(\mu_{\alpha_1}, \sigma_{\alpha_1})$$

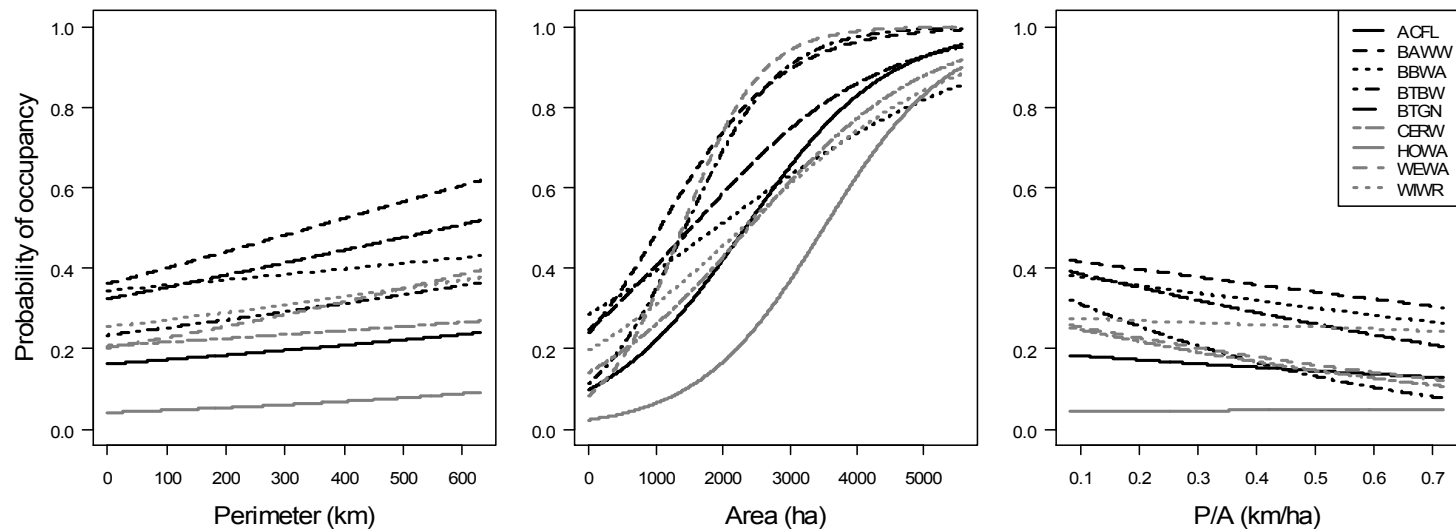
where

μ_{α_1} = mean effect of forest fragment perimeter on species occupancy

σ_{α_1} = standard deviation of perimeter effect across species

Multi-species occupancy model

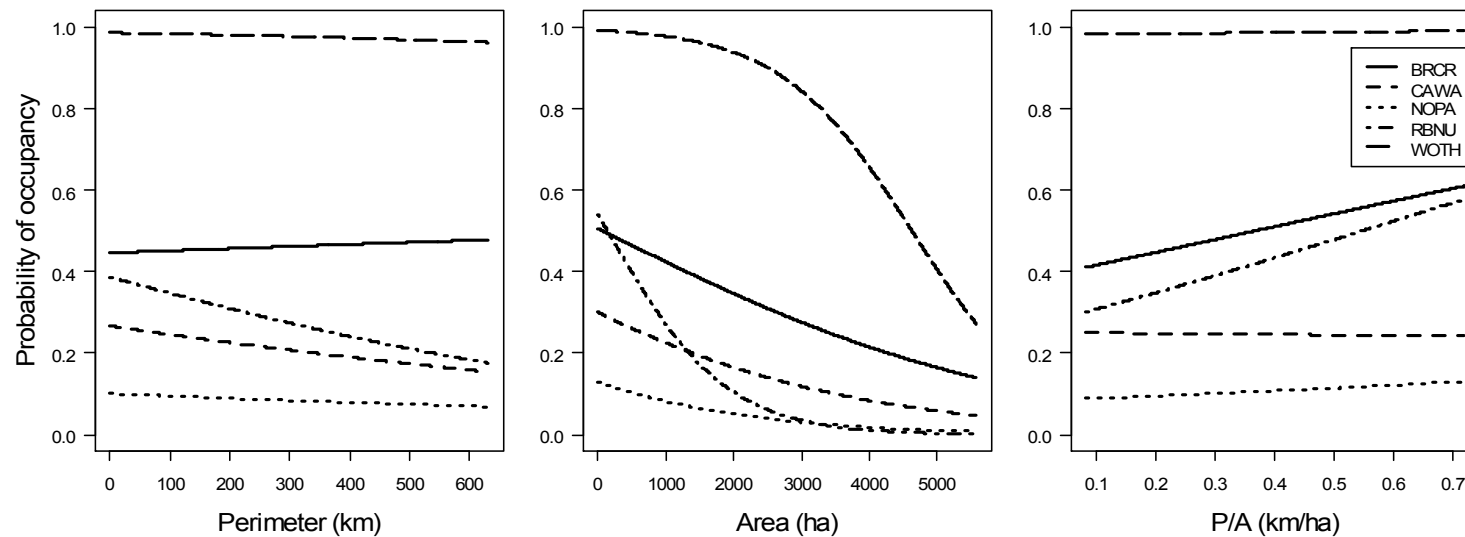
Results



Mean marginal probabilities of occupancy for nine forest interior breeding bird species

Multi-species occupancy model

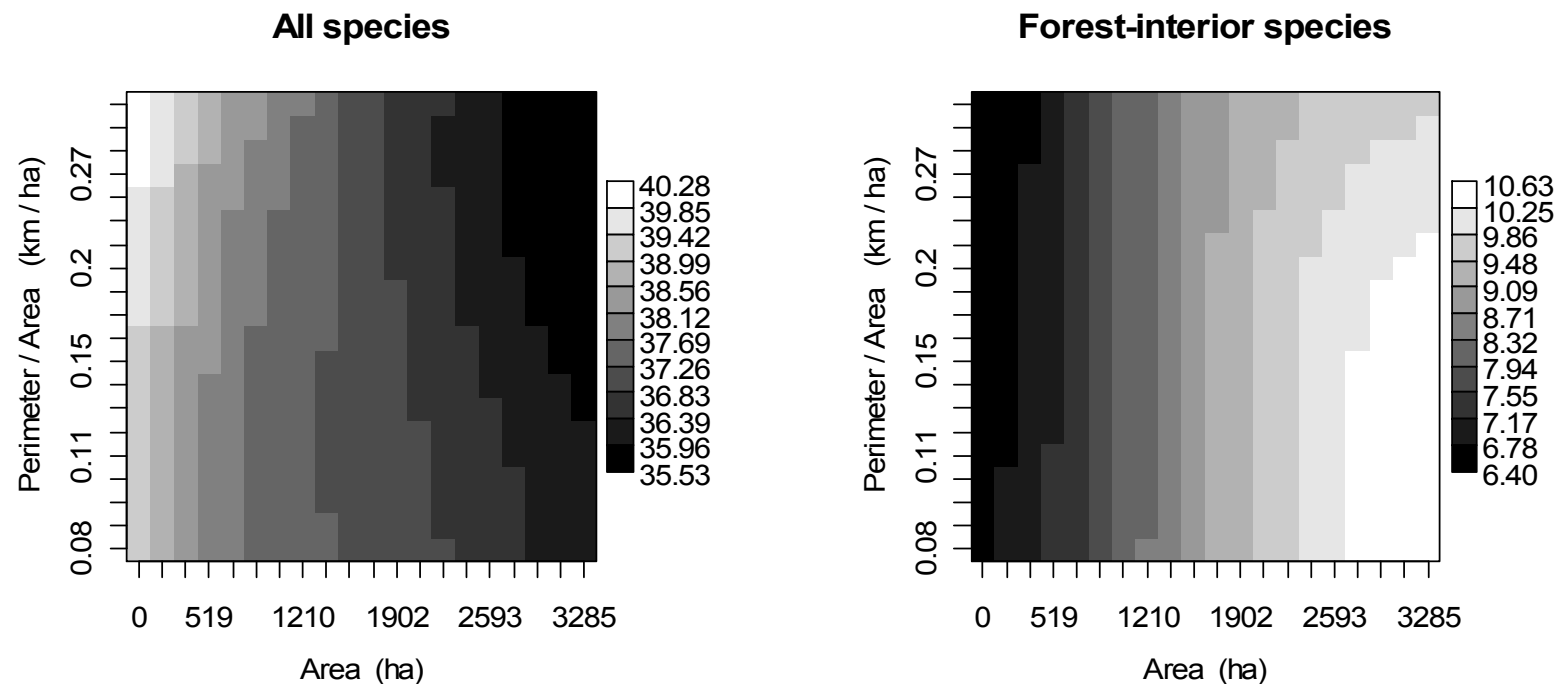
Results



Mean marginal probabilities of occupancy for five other forest interior breeding bird species

Multi-species occupancy model

Results



Total estimated species richness (left) based on area and P/A and estimated assemblage richness of forest interior species

Multi-species occupancy model

Example: Conclusions

- Species all respond differently to forest fragmentation -> if we had lumped all forest interior species together, we would have missed important species differences
- Richness was highest in small forest fragments with high P/A values -> edges promote richness
- BUT richness for forest interior species (i.e., those ones with the greatest conservation values) was highest in the largest forest fragments with minimal edge habitat.

Multi-species occupancy models

Homework: Estimating bird occupancy using a multi-species model