Bayesian occupancy modeling with the spOccupancy R package

Jeff Doser

Michigan State University

July 27, 2022

Tel Aviv University Seminar

@jeffdoser18

Overview

- Overview of occupancy modeling and spatial autocorrelation
- spOccupancy functionality
- spOccupancy syntax and example: single-species occupancy model
- Additional resources

Motivation

- Species distribution modeling
  - Where do species occur and how does this change over time?
  - What drives species distributions?
- Two key complexities when modeling species distributions
  - Imperfect detection
  - Spatial autocorrelation

Imperfect Detection

How do we account for imperfect detection?

- Occupancy modeling
- Basic idea: perform multiple surveys (i.e., visits) at each site
- Multiple visits give information on detection probability
- Allows us to separately estimate occupancy probability from detection probability

Detection-nondetection matrix

Site	Survey 1	Survey 2	Survey 3	Survey 4
1	1	0	0	1
2	0	0	0	0
3	1	1	0	NA
4	1	NA	0	NA
5	0	1	1	1
6	0	0	0	1

Occupancy model: what is it?

- Two distinct sub-models
  - Model occupancy probability as a function of site-level covariates
  - Model detection probability as a function of site and/or survey-level covariates
    - Can only detect a species if it truly occupies a site
    - Detection probability is modeled "conditional" on true occupancy

Occupancy model

Occupancy (ecological) sub-model

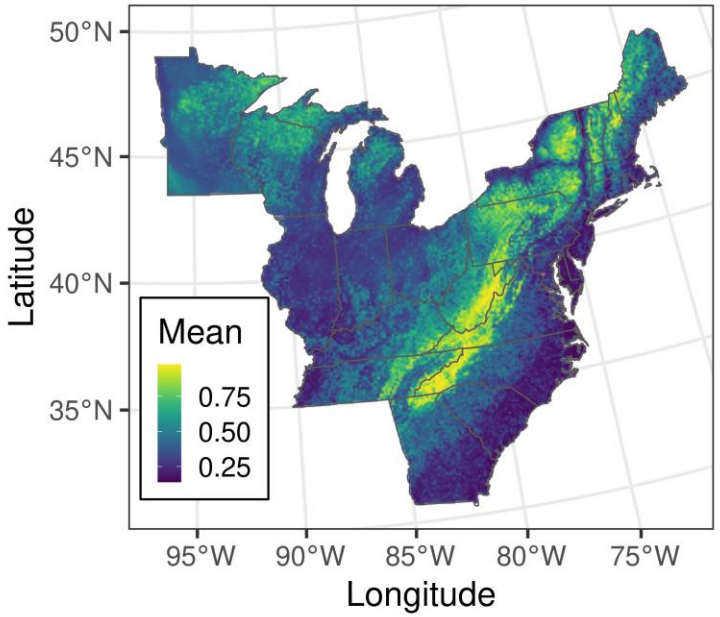
$$z_j \sim \text{Bernoulli}(\psi_j)$$
$$\text{logit}(\psi_j) = \beta_1 + \beta_2 \cdot X_{2,j} + \cdots + \beta_r \cdot X_{r,j}$$

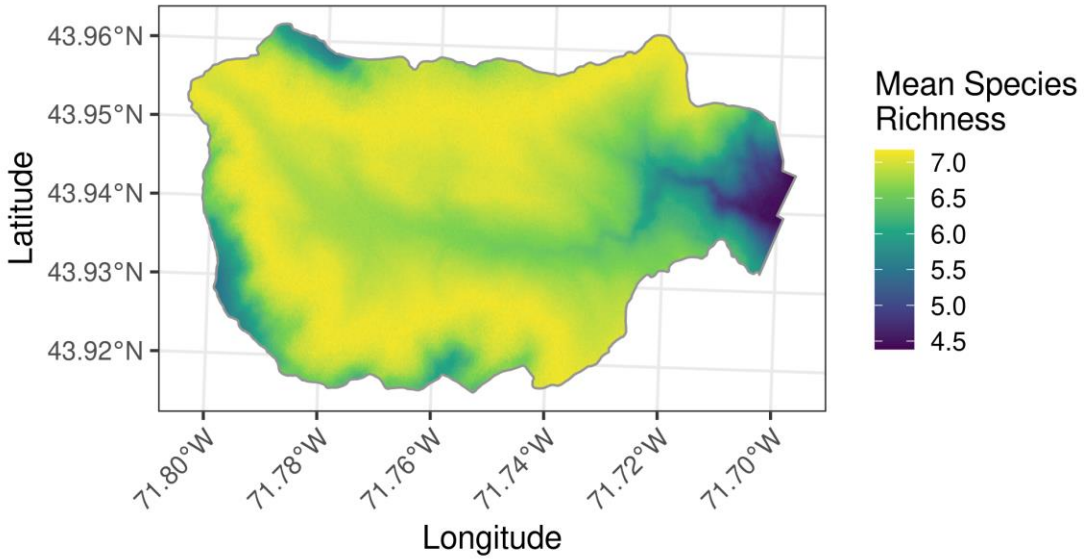
Detection (observation) sub-model

$$y_{j,k} \sim \text{Bernoulli}(p_{j,k} \cdot z_j)$$
$$\text{logit}(p_{j,k}) = \alpha_1 + \alpha_2 \cdot V_{2,j,k} + \cdots + \alpha_r \cdot V_{r,j,k}$$

Spatial autocorrelation

- Things closer together in space tend to be more similar than things further apart
- What leads to spatial autocorrelation in species distributions?
  - Environmental drivers, habitat requirements
  - Biotic factors (dispersal, conspecific attraction)
- Usual approach: explain spatial variation in species distributions with covariates (e.g., forest cover, temperature)





Doser et al. (2022) MEE

Residual spatial autocorrelation

- Spatial correlation in data *after* including spatial covariates
- Often arises from missing/unavailable covariates
- Can lead to bias if unaddressed
- Account for using spatial random effects
  - Each site has a local adjustment in occupancy probability
  - The local adjustments are given a spatial structure
  - Estimated parameters: spatial variance and spatial range

Spatial occupancy model

Occupancy (ecological) sub-model

$$z_j \sim \text{Bernoulli}(\psi_j)$$
$$\text{logit}(\psi_j) = \beta_1 + \beta_2 \cdot X_{2,j} + \cdots + \beta_r \cdot X_{r,j} + \text{w}_j$$
$$\text{w}_j \sim \text{Normal}(0, \boldsymbol{\Sigma})$$

Detection (observation) sub-model

$$y_{k,j} \sim \text{Bernoulli}(p_{j,k} \cdot z_j)$$
$$\text{logit}(p_{j,k}) = \alpha_1 + \alpha_2 \cdot V_{2,j,k} + \cdots + \beta_r \cdot V_{r,j,k}$$

spOccupancy

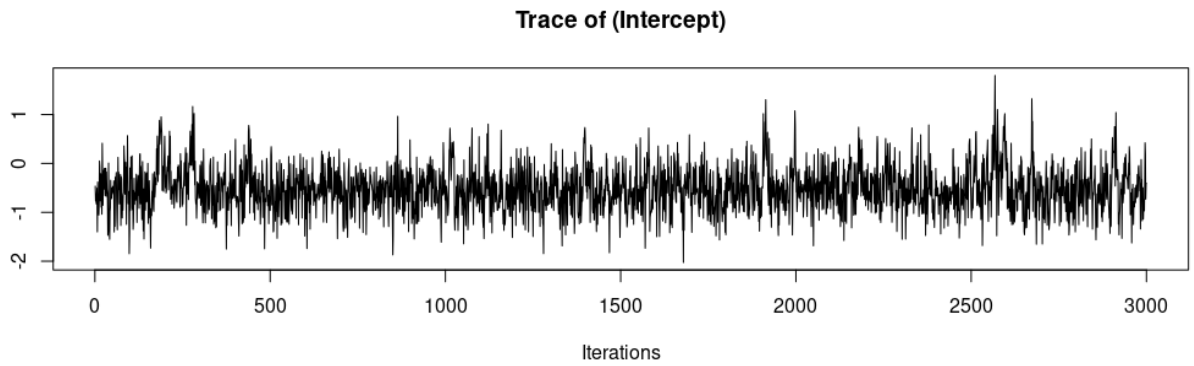
- Designed to fit a variety of Bayesian occupancy models
- Options to efficiently accommodate spatial autocorrelation
- Workflow completely in R using standard model syntax (no Bayesian programming languages necessary)
- Key functionality:
  - Single-species models
  - Multi-species models
  - Data integration
  - Multi-season (spatio-temporal) models (hot off the press!)

Why Bayesian for occupancy modeling?

- Interpretation
- More flexible to accommodate spatial autocorrelation
- Easy to extend to multi-species frameworks/integrate multiple data sources
- Fully propagate uncertainty in all estimates (and derived quantities)

Bayesian computation: what to know

- Markov Chain Monte Carlo (MCMC)
- MCMC "chains" eventually converge to a posterior distribution (what we use to compute means, credible intervals, etc)
  - Assess convergence by running multiple chains with different starting values



Bayesian computation: what to specify

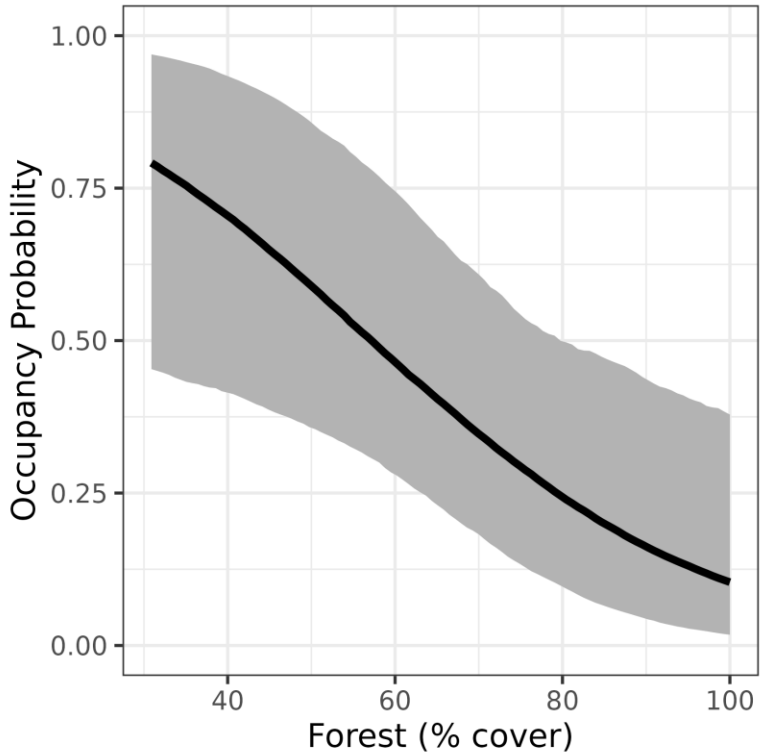
- What do you need to specify?
  - Priors
  - Initial values
  - Number of samples/iterations
  - Burn-in: initial part of the chain that we throw away
  - Thinning rate: how often do you want to save a sample?

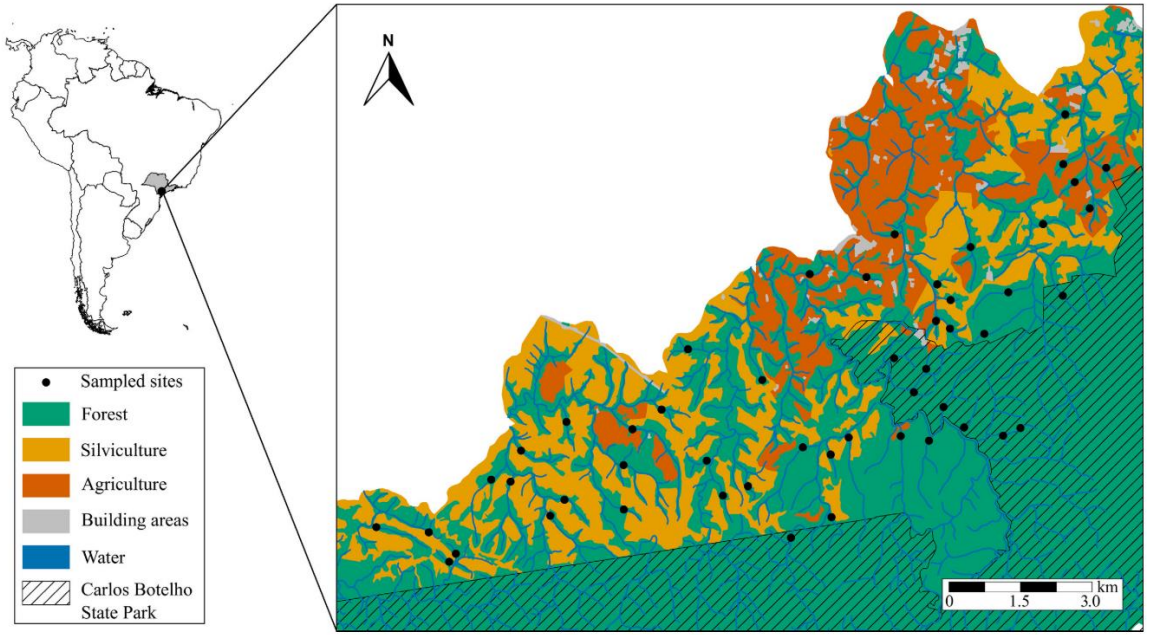
spOccupancy workflow

1. Data simulation/prep
2. Model fitting
3. Model validation
4. Model comparison
5. Posterior summaries
6. Prediction

Single-species occupancy model example

- Data from Ribeiro Jr. Et al (2018) *Eco Apps*
- 50 sites along a gradient of landscape characteristics
- Focus on the tropical frog species *Crossodactylus caramaschii*





Ribeiro Jr. et al. (2018) *Eco Apps*

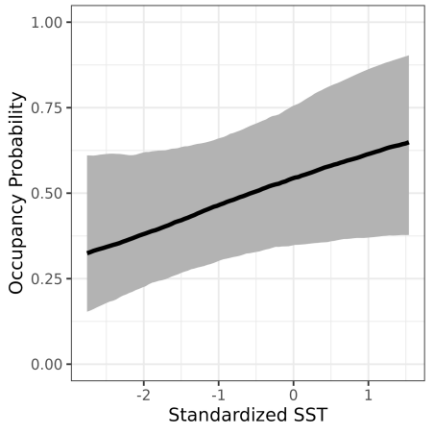
Additional examples

Multi-species occupancy model



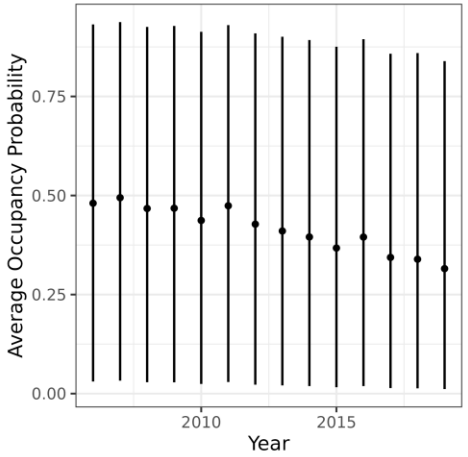
Amphibian community from Ribeiro Jr et al (2018) Eco Apps

Integrated occupancy model



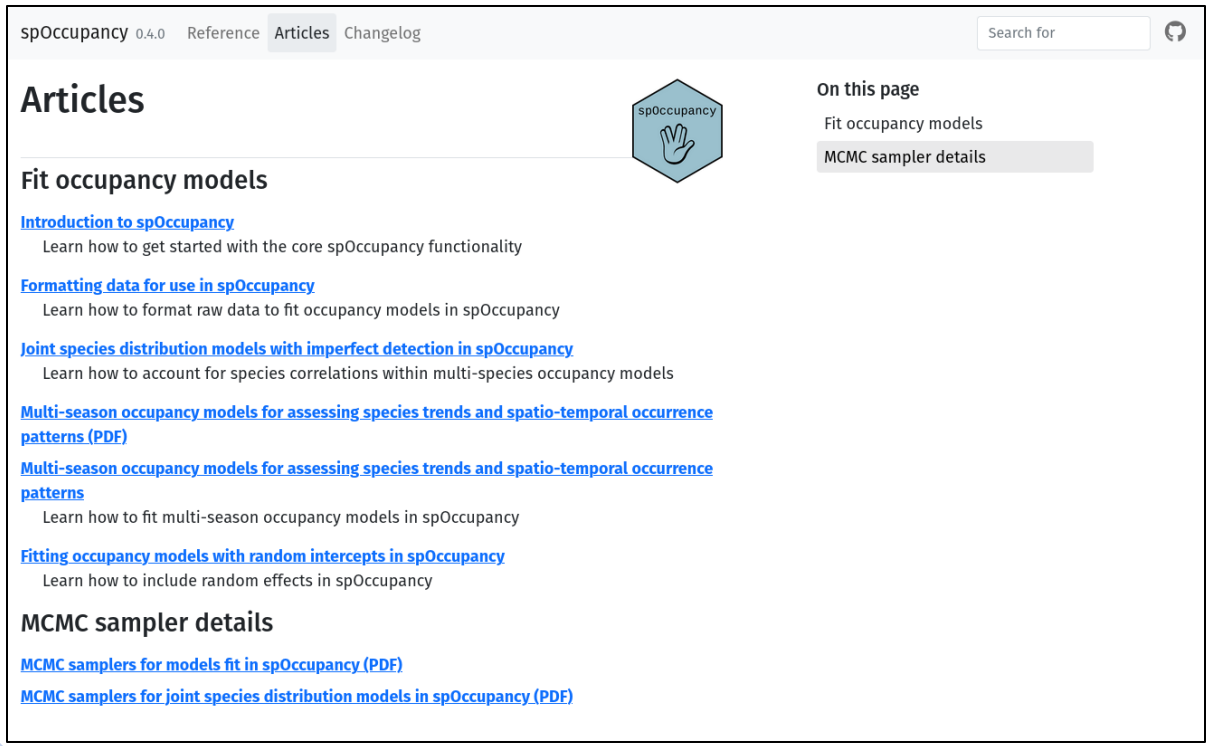
Bottlenose dolphin data from Lauret et al. (2021) Ecology

Multi-season occupancy model

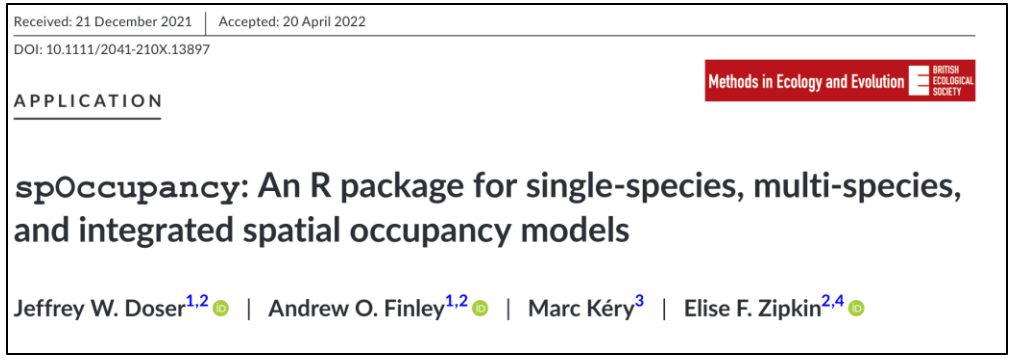


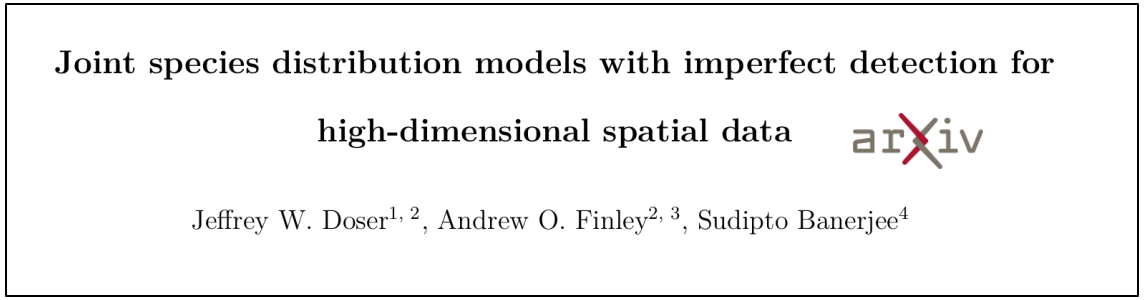
Eastern wood pewee data from Doser et al. (2021) Eco Apps

spOccupancy



- Package website
  - https://www.jeffdoser.com/files/spoccupancy-web/
- GitHub development page
  - https://github.com/doserjef/spOccupancy/
- MEE intro paper
- arXiv preprint
- @jeffdoser18
- Email: doserjef@msu.edu





Acknowledgements

Andy Finley

Elise Zipkin

Marc Kéry

Sudipto Banerjee

MICHIGAN STATE UNIVERSITY

NSF

Thank you!