Spatial occupancy models with the spOccupancy R package

Jeff Doser and Andy Finley

Michigan State University

October 3, 2022

ESA Statistical Methods Seminar Series

Overview

- Overview of occupancy modeling and spatial modeling
- spOccupancy functionality
- spOccupancy syntax and example: single-species and multi-species spatial occupancy models
- Q&A

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

Single-species occupancy model

$$j = 1, \ldots, J \text{ (site)}$$
$$k = 1, \ldots, K_j \text{ (replicate)}$$

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

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

Multi-species occupancy models

- Two flavors in spOccupancy:
  - Dorazio-Royle multi-species (community) occupancy models
    - Spatial and non-spatial
  - Joint species distribution models with imperfect detection
    - Spatial and non-spatial
    - Account for species correlations using factor modeling (Doser, Finley, Banerjee 2022)

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

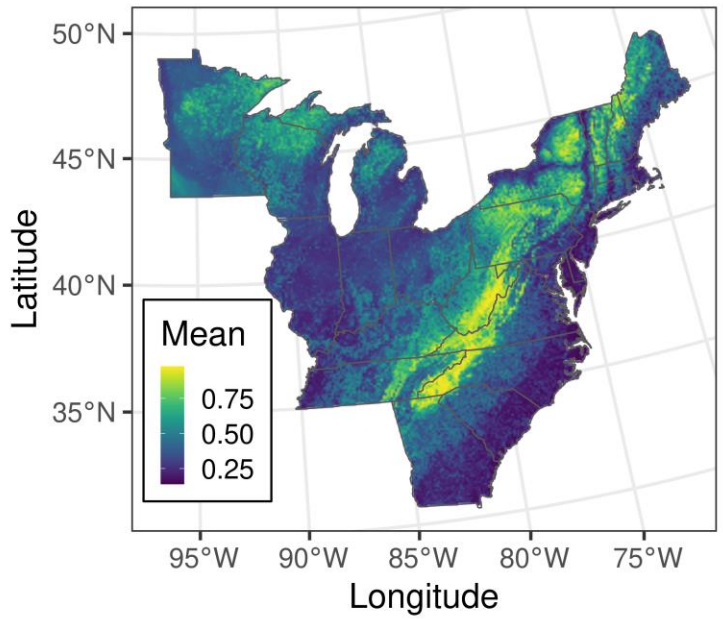
Site	Survey	Survey	Survey	Survey 4
				0
				0
				NA
				NA
				1
				1

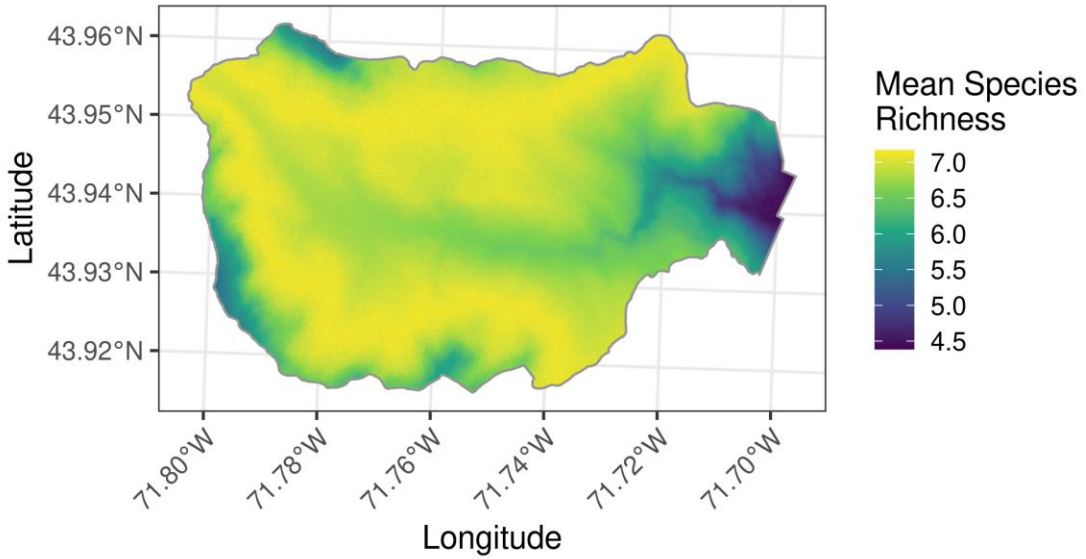
Site	Survey	Survey	Survey	Survey 4
				1
				01
				NA
				NA
				1
				1

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

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)
- Initial approach: attempt to explain spatial variation in species distributions with covariates (e.g., forest cover, temperature, elevation)





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 decay
  - Caution: spatial confounding (Hanks et al. 2015)

Single-species spatial occupancy model

$$j = 1, \ldots, J \text{ (site)}$$
$$k = 1, \ldots, K_j \text{ (replicate)}$$

Occupancy (ecological) sub-model

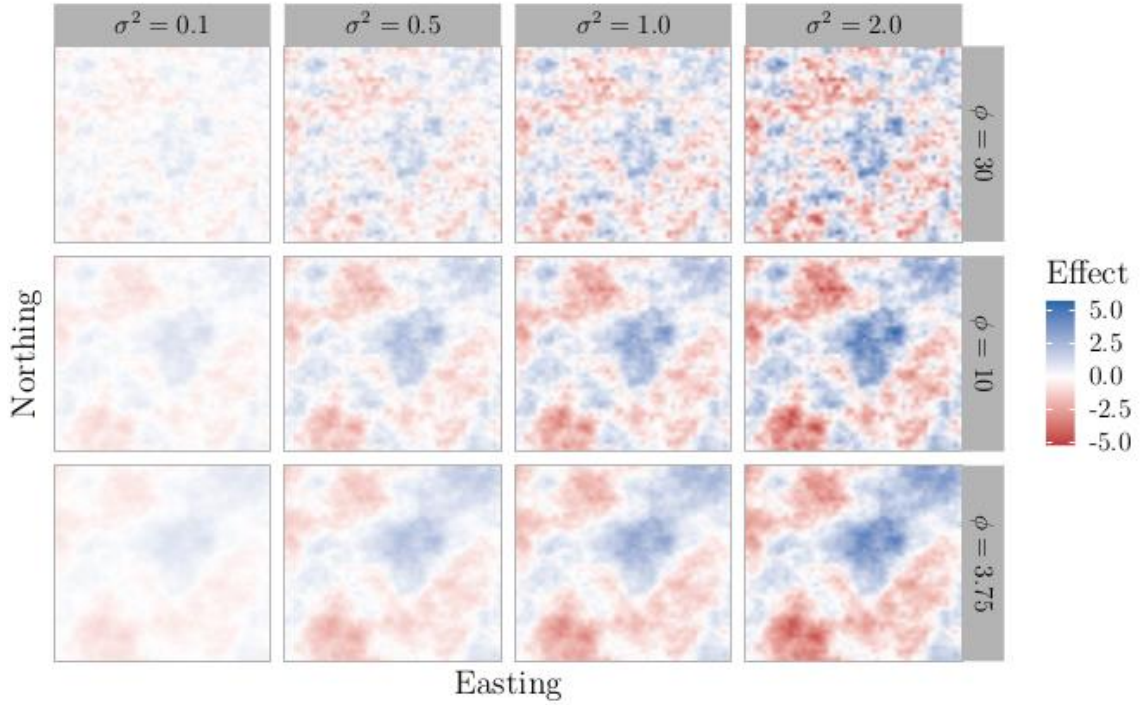
$$z_j \sim \text{Bernoulli}(\psi_j)$$
$$\text{logit}(\psi_j) = \beta_1 + \beta_2 \cdot X_{2,j} + \cdots + \beta_r \cdot X_{r,j} + \text{w}_j$$
$$\text{w}_j \sim \text{Normal}(0, \boldsymbol{\Sigma})$$

Detection (observation) sub-model

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

Gaussian processes

- "Gold standard" for modeling spatial data
- Spatial correlation function determines structure
- Downside: computationally intensive for big data
- Alternative: Nearest Neighbor Gaussian Processes (Datta et al. 2016, Doser et al. 2022)



spOccupancy

- Designed to fit a variety of Bayesian occupancy models
- Efficiently accommodates spatial autocorrelation
- Workflow completely in R using standard model syntax (no Bayesian programming languages necessary)
- Key functionality:
  - Single-species models
  - Multi-species models with options to account for species correlations
  - Data integration
  - Multi-season (spatio-temporal) models

Why Bayesian for occupancy modeling?

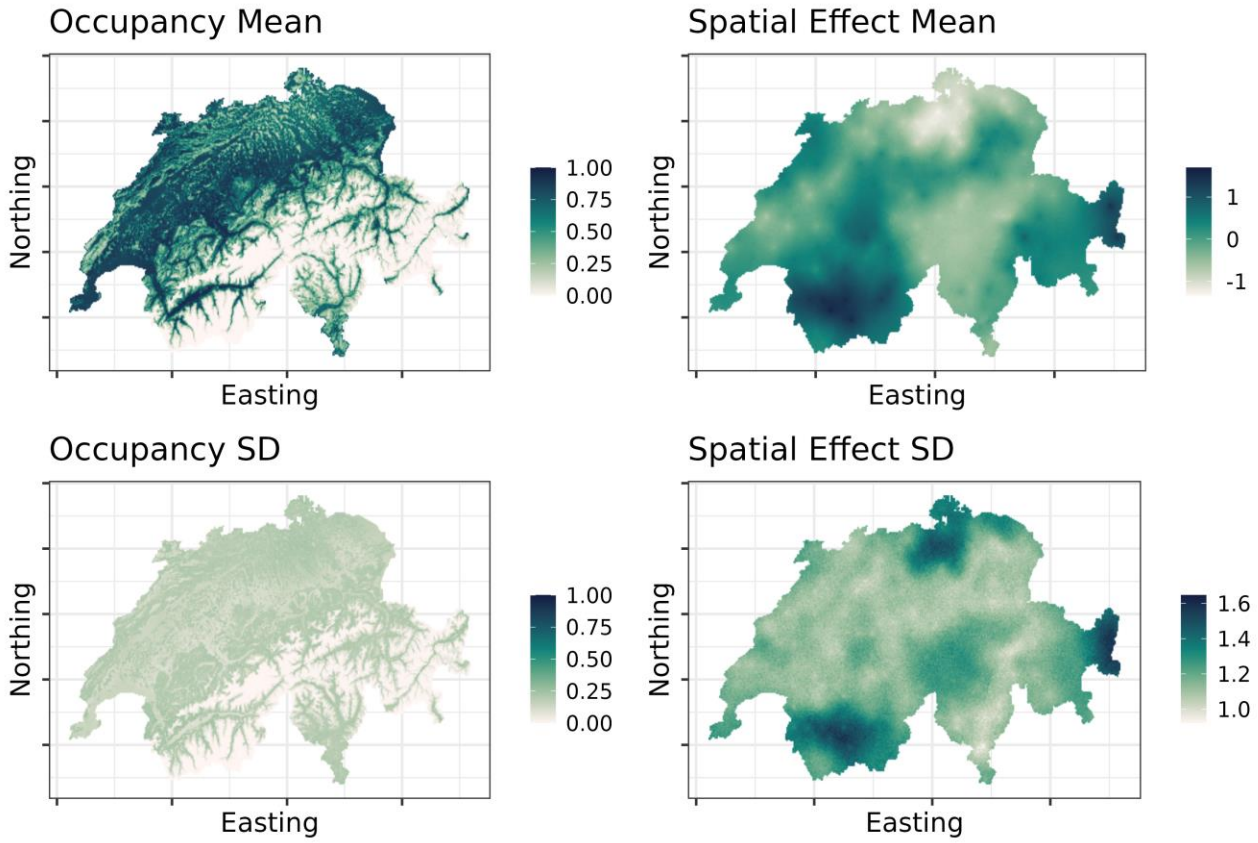
- Interpretation
- More flexible to accommodate spatial autocorrelation
- Easy to extend to multi-species frameworks/integrate multiple data sources
- Fully propagate uncertainty in all estimates (and derived quantities)

spOccupancy workflow

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

Example dataset: Swiss MHB Survey

- Data from the Switzerland Breeding Bird Survey in 2014
- 266 survey locations distributed throughout Switzerland
- Single species (European Goldfinch) and multi-species models (20 species)



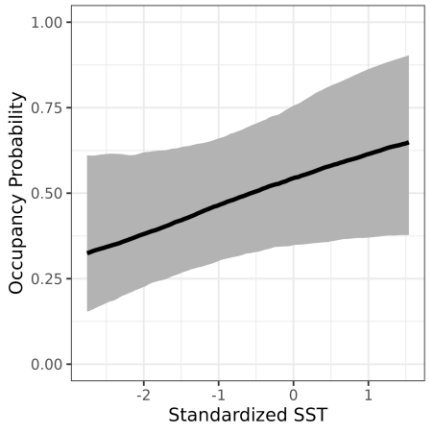
Additional examples

Single-species and multi-species occupancy models



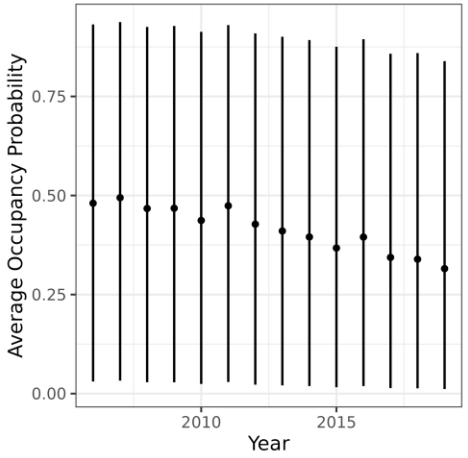
Amphibian community from Ribeiro Jr et al (2018) Eco Apps

Integrated occupancy model



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

Multi-season occupancy model

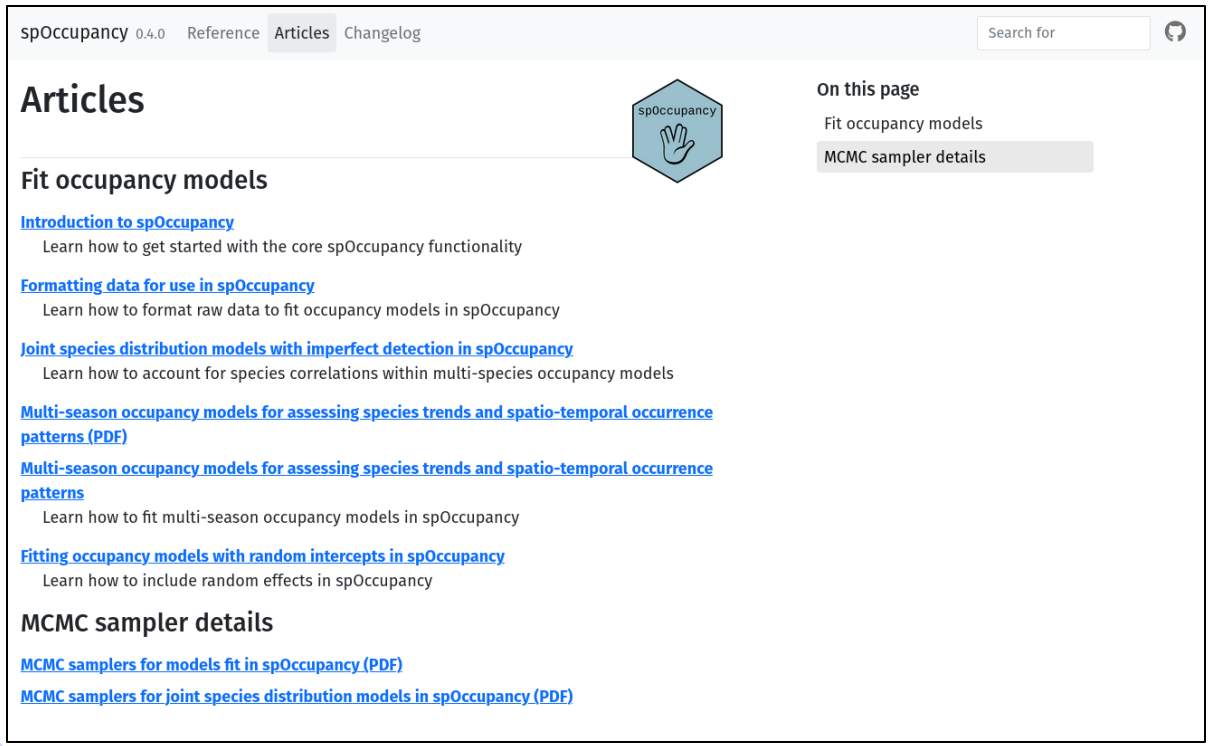


Eastern wood pewee data from Doser et al. (2021) Eco Apps

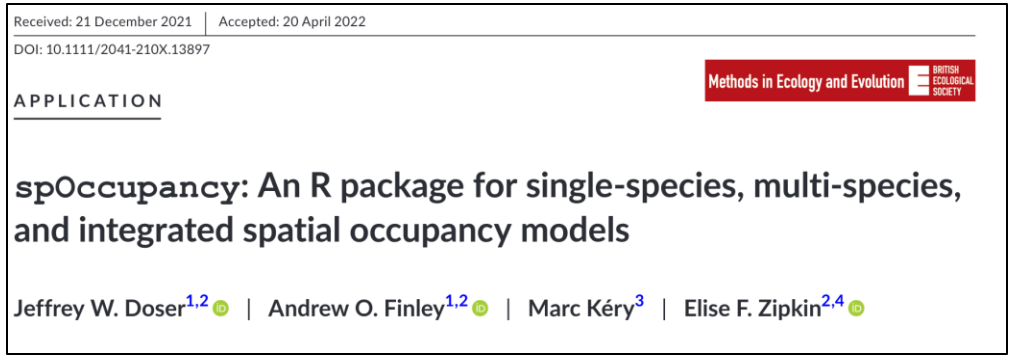
Coming soon...

- Spatially-varying coefficient occupancy models
- Multi-species spatio-temporal occupancy models
- Multi-species integrated occupancy models

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


Received: 21 December 2021 | Accepted: 20 April 2022
DOI: 10.1111/2041-210X.13897
Methods in Ecology and Evolution
APPLICATION

spOccupancy: An R package for single-species, multi-species, and integrated spatial occupancy models

Jeffrey W. Doser[1,2] | Andrew O. Finley[1,2] | Marc Kéry[3] | Elise F. Zipkin[2,4]

Joint species distribution models with imperfect detection for high-dimensional spatial data

Jeffrey W. Doser[1, 2], Andrew O. Finley[2, 3], Sudipto Banerjee[4]

Acknowledgements


Elise Zipkin


Marc Kéry


Sudipto Banerjee

MICHIGAN STATE UNIVERSITY

NSF

Thank you!