

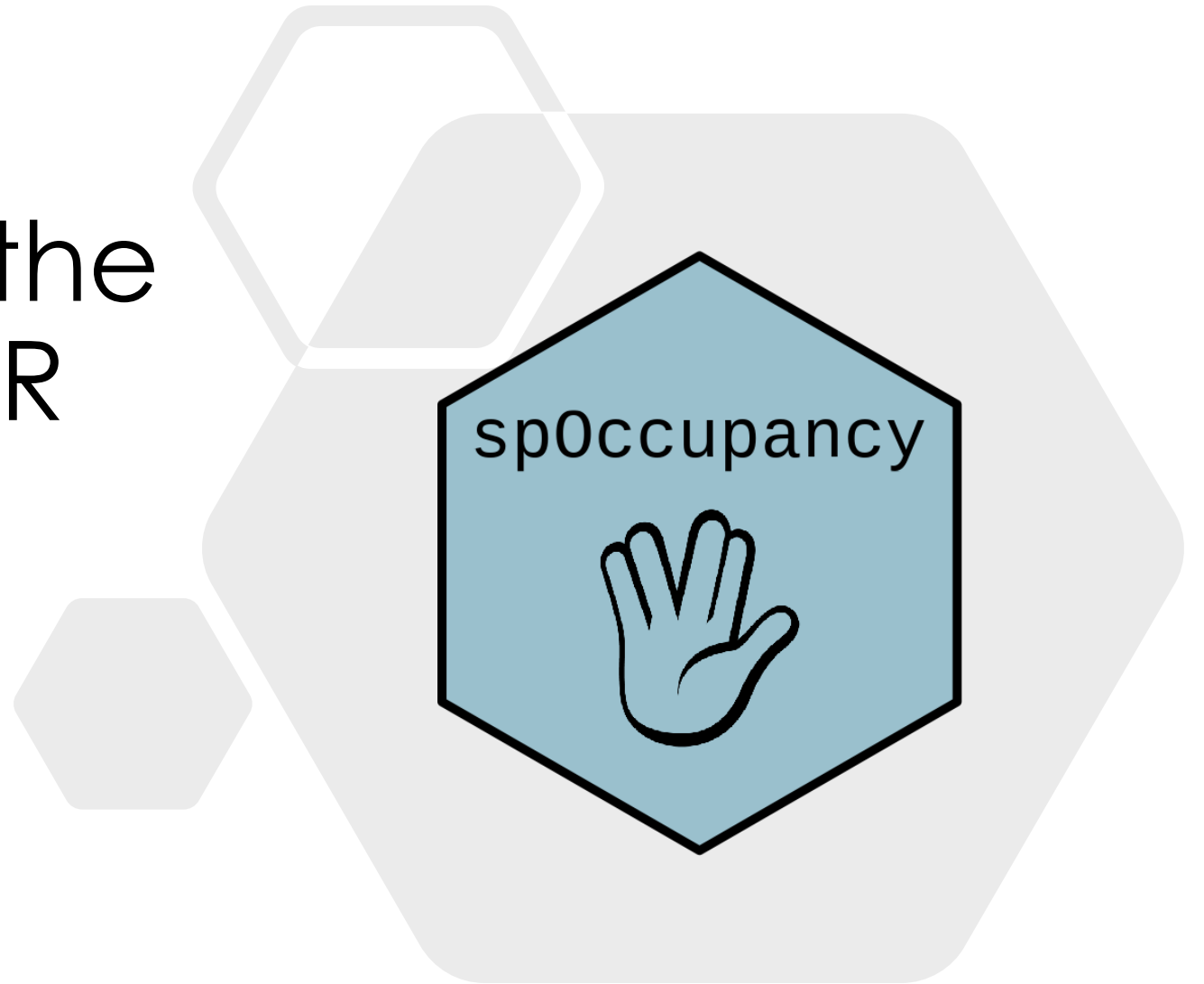
Bayesian occupancy modeling with the spOccupancy R package

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

README.md

Bayesian occupancy modeling with the spOccupancy R package

[Jeffrey W. Doser](#)

Department of Integrative Biology, Michigan State University

About

Occupancy modeling is a common approach to assess species distribution patterns across space and/or time while explicitly accounting for false absences in detection-nondetection data. Numerous extensions of the basic single-species occupancy model exist to model multiple species, spatial autocorrelation and to integrate multiple data types. This presentation discusses [spOccupancy](#), an R package designed to fit a variety of Bayesian single-species and multi-species occupancy models. I first give a brief introduction of occupancy modeling as a robust form of species distribution model as well as spatial autocorrelation and how it arises in detection-nondetection data. I then introduce the `spOccupancy` package and detail how to fit single-species spatial and non-spatial occupancy models. In this repository, I provide examples of three additional forms of occupancy models fit by `spOccupancy`: (1) multi-species models, (2) integrated occupancy models, and (3) multi-season (spatio-temporal) occupancy models. The repository contains the presentation slides, the example data sets, and the R scripts for the four examples.


Installing spOccupancy

spOccupancy can be installed from CRAN using `install.packages("spOccupancy")`.

Installing additional R packages

In the examples, we will use three additional packages for summarizing our results: [coda](#), [MCMCvis](#), and [ggplot2](#). These packages can be installed from CRAN as follows:

```
install.packages(c("coda", "MCMCvis", "ggplot2"))
```



Releases

No releases published
[Create a new release](#)

Packages

No packages published
[Publish your first package](#)

Languages

R 100.0%

https://github.com/doserjef/spOccupancy_examples

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

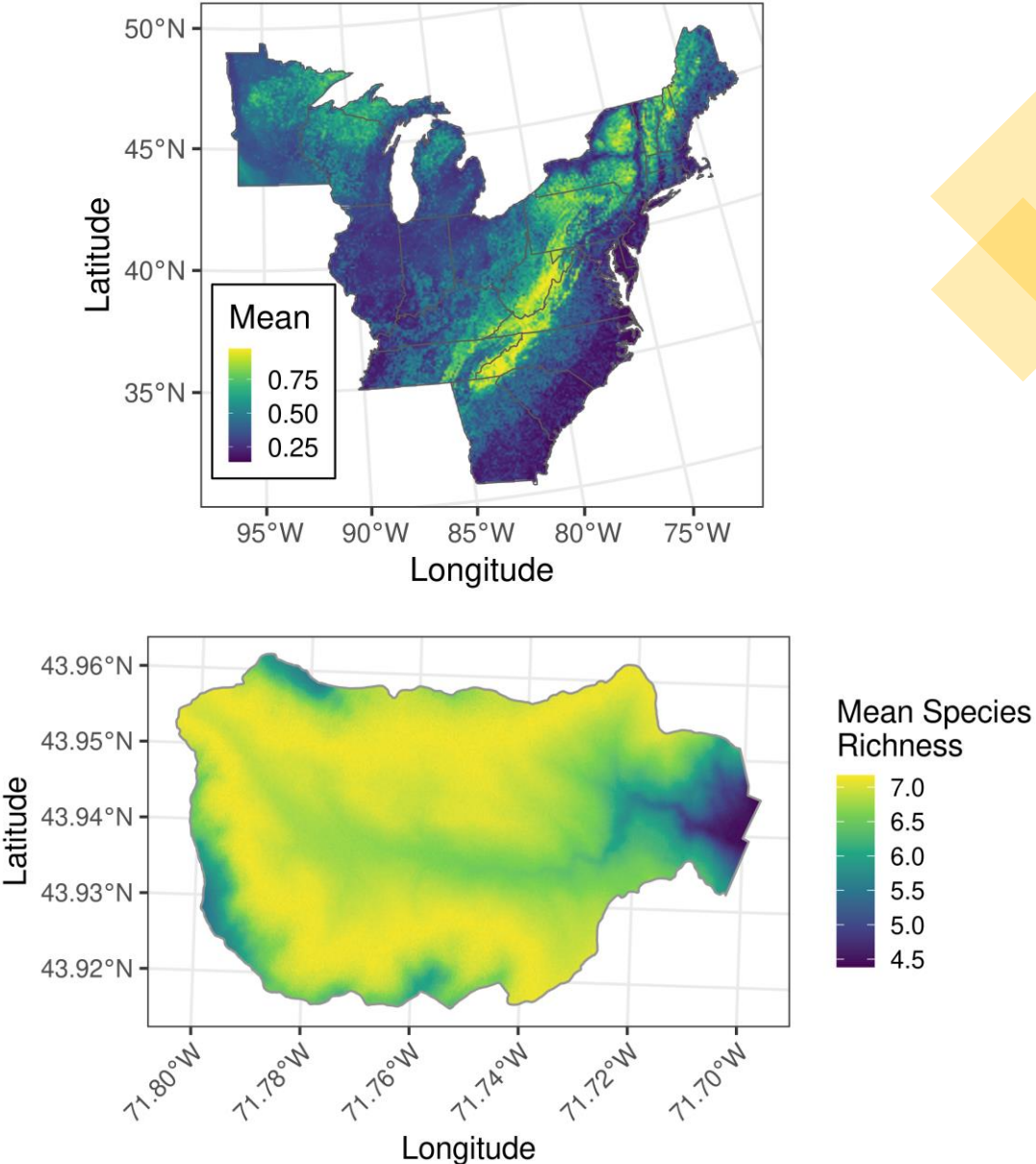
$$z_j \sim \text{Bernoulli}(\psi_j)$$
$$\text{logit}(\psi_j) = \beta_1 + \beta_2 \cdot X_{2,j} + \dots + \beta_r \cdot X_{r,j}$$

Detection (observation) sub-model

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

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)



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

$$\begin{aligned}z_j &\sim \text{Bernoulli}(\psi_j) \\ \text{logit}(\psi_j) &= \beta_1 + \beta_2 \cdot X_{2,j} + \cdots + \beta_r \cdot X_{r,j} + w_j \\ w_j &\sim \text{Normal}(0, \Sigma)\end{aligned}$$

Detection (observation) sub-model

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

spOccupancy



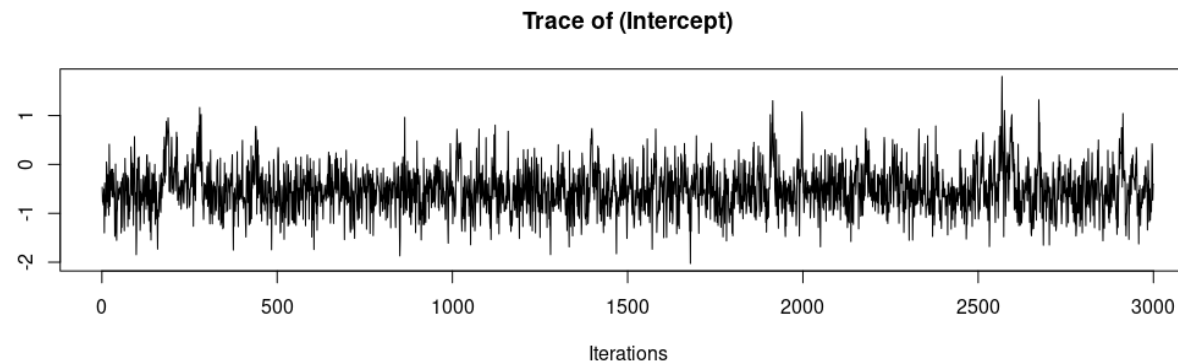
- Designed to fit a variety of Bayesian occupancy models
- Options to efficiently accommodate spatial autocorrelation
- Workflow completely in R using standard model syntax (no Bayesian programming languages necessary)
- Key functionality:
 - Single-species models
 - Multi-species models
 - Data integration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

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

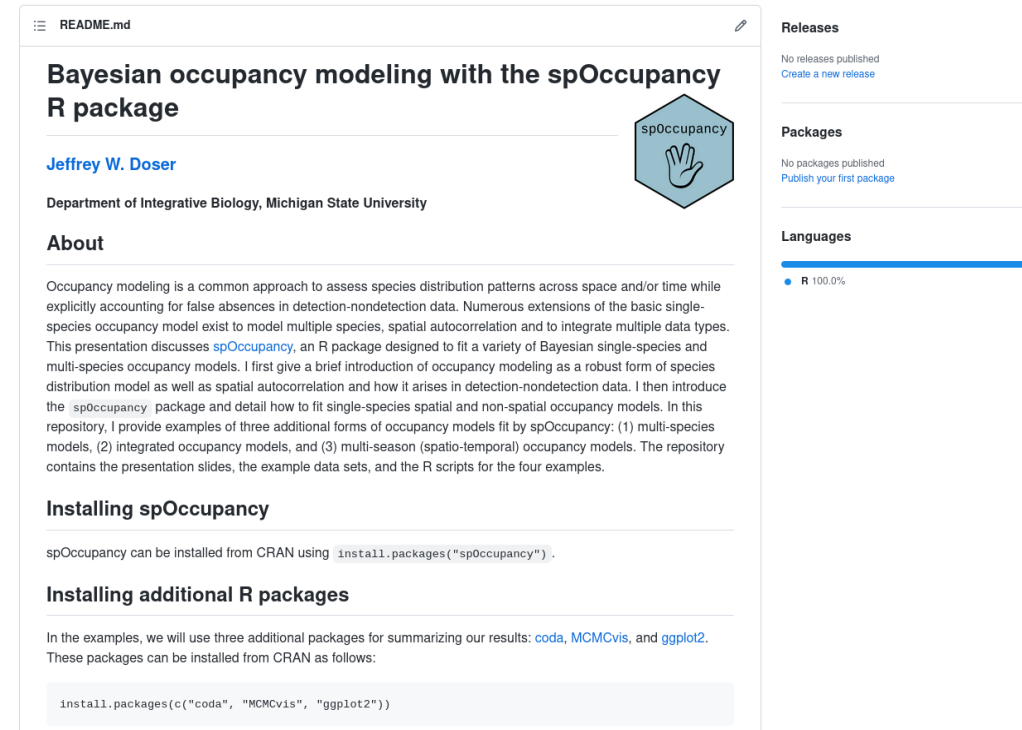


Bayesian computation: what to specify

- 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

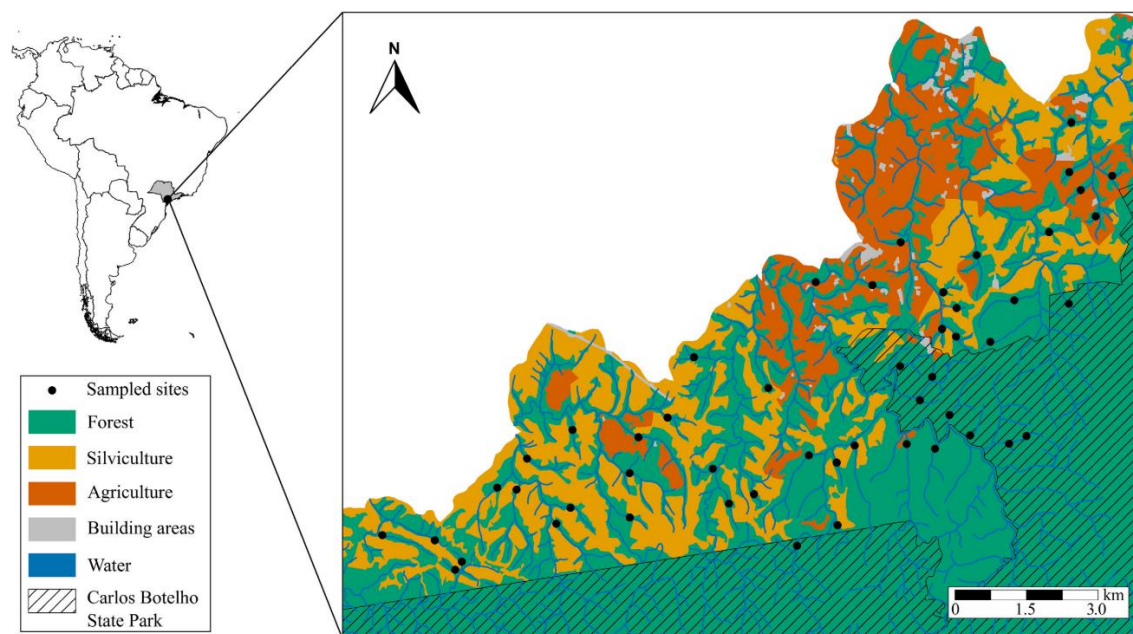
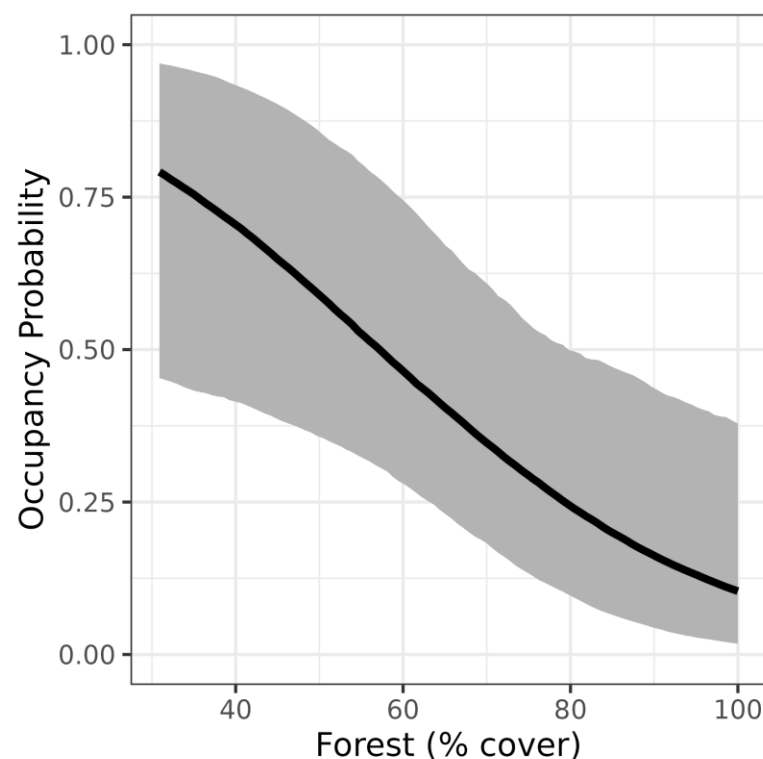


The screenshot shows the GitHub README for the **spOccupancy** R package. The title is "Bayesian occupancy modeling with the spOccupancy R package" by Jeffrey W. Doser. The author's affiliation is the Department of Integrative Biology, Michigan State University. The "About" section describes the package's purpose in modeling species distribution patterns. The "Installing spOccupancy" section provides the CRAN installation command: `install.packages("spOccupancy")`. The "Installing additional R packages" section lists `coda`, `MCMCvis`, and `ggplot2` as dependencies and provides the installation command: `install.packages(c("coda", "MCMCvis", "ggplot2"))`. On the right sidebar, the "Releases" section shows "No releases published" with a link to "Create a new release". The "Packages" section shows "No packages published" with a link to "Publish your first package". The "Languages" section shows a progress bar for R at 100.0%.

https://github.com/doserjef/spOccupancy_examples

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*

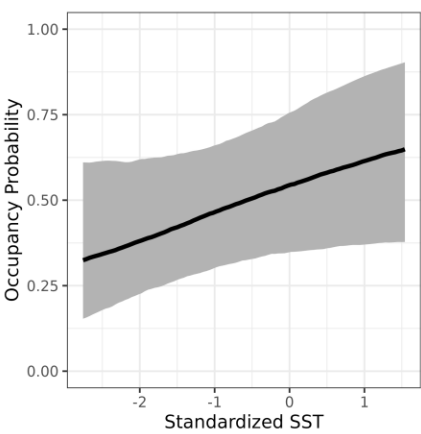
Additional examples

Multi-species occupancy model



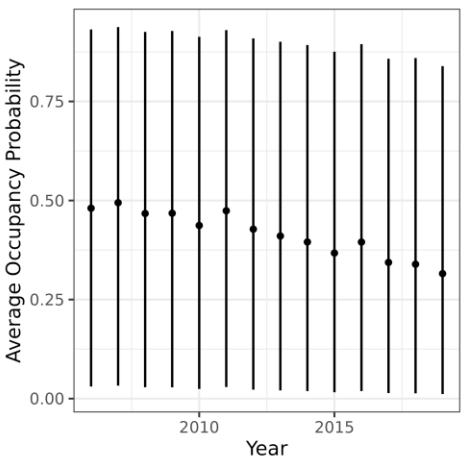
Amphibian community from Ribeiro Jr et al (2018) Eco Apps

Integrated occupancy model



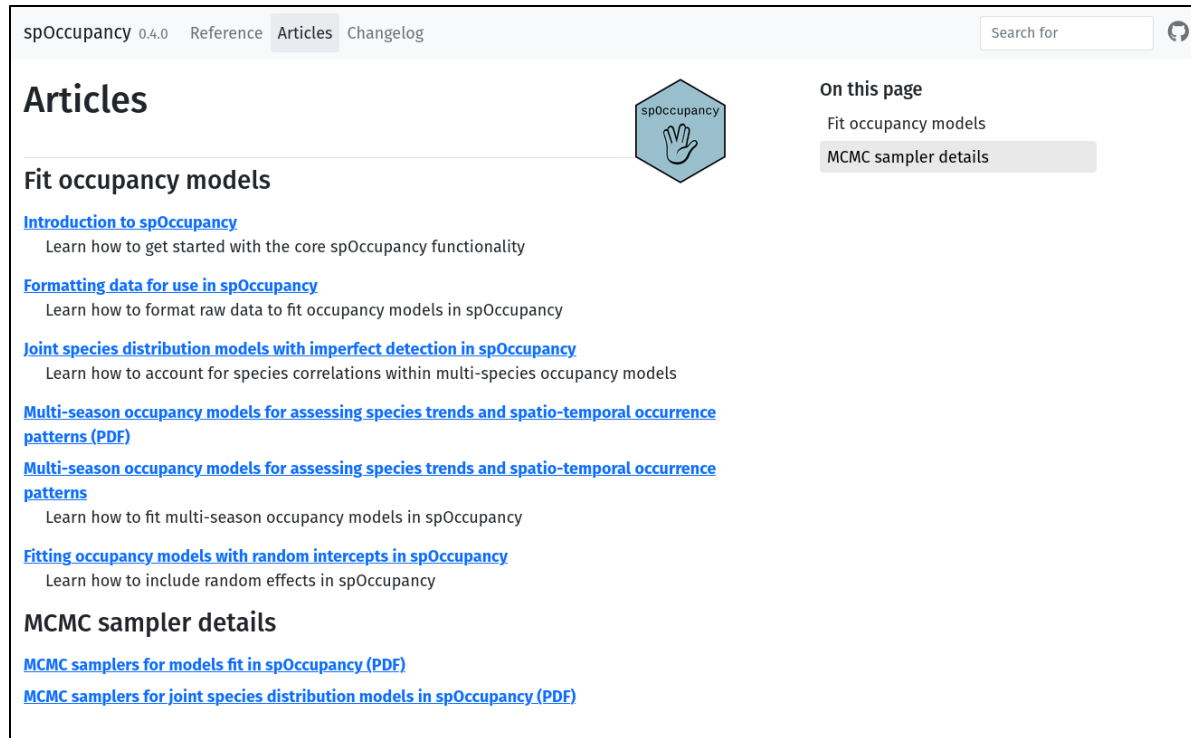
Bottlenose dolphin data from Lauret et al. (2021) Ecology

Multi-season occupancy model



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

spOccupancy




The screenshot shows the spOccupancy website interface. At the top, there are navigation links: "spOccupancy 0.4.0", "Reference", "Articles", and "Changelog". A search bar is located on the right. The main heading is "Articles". Below it, there is a section titled "Fit occupancy models" with a list of links: "Introduction to spOccupancy", "Formatting data for use in spOccupancy", "Joint species distribution models with imperfect detection in spOccupancy", "Multi-season occupancy models for assessing species trends and spatio-temporal occurrence patterns (PDF)", and "Multi-season occupancy models for assessing species trends and spatio-temporal occurrence patterns". Each link is followed by a brief description. To the right of the articles, there is a sidebar titled "On this page" with links to "Fit occupancy models" and "MCMC sampler details".

- 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



The image shows the cover of the paper "spOccupancy: An R package for single-species, multi-species, and integrated spatial occupancy models" published in the journal "Methods in Ecology and Evolution". The cover includes the journal logo, the title, and the authors: Jeffrey W. Doser^{1,2}, Andrew O. Finley^{1,2}, Marc Kéry³, and Elise F. Zipkin^{2,4}. It also displays the DOI: 10.1111/2041-210X.13897 and the dates: Received: 21 December 2021, Accepted: 20 April 2022.

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

Jeffrey W. Doser^{1, 2}, Andrew O. Finley^{2, 3}, Sudipto Banerjee⁴

Acknowledgements



Andy Finley



Elise Zipkin



Marc Kéry



Sudipto Banerjee



Thank you!