Hierarchical spatial modelling for applied population and community ecology

Jeffrey W. Doser, Marc Kéry, Gesa von Hirschheydt

24-27 June 2024

Introduction to occupancy models

Jeffrey W. Doser

24-27 June 2024

Ecological Motivation

- Conservation management often requires information on species distributions across space and time

Ecological Motivation

- Conservation management often requires information on species distributions across space and time
- Examples
  - Species occurrence in protected vs. unprotected areas
  - Management effects on a community of species
  - Trends in species occurrence over time as a measure of population change

Ecological Motivation

- Conservation management often requires information on species distributions across space and time
- Examples
  - Species occurrence in protected vs. unprotected areas
  - Management effects on a community of species
  - Trends in species occurrence over time as a measure of population change
- Two important complexities when modelling species distributions:
  - Imperfect detection

Ecological Motivation

- Conservation management often requires information on species distributions across space and time
- Examples
  - Species occurrence in protected vs. unprotected areas
  - Management effects on a community of species
  - Trends in species occurrence over time as a measure of population change
- Two important complexities when modelling species distributions:
  - Imperfect detection
  - Spatial autocorrelation

Ecological Motivation

- Conservation management often requires information on species distributions across space and time
- Examples
  - Species occurrence in protected vs. unprotected areas
  - Management effects on a community of species
  - Trends in species occurrence over time as a measure of population change
- Two important complexities when modelling species distributions:
  - **Imperfect detection**
  - Spatial autocorrelation

What data do we use to quantify species distributions?

What data do we use to quantify species distributions?

- Presence-only data: locations where species occur

What data do we use to quantify species distributions?

- Presence-only data: locations where species occur
- Presence-absence data: data takes value 1 if a species is present at some location and/or time point and 0 if the species is absent

What data do we use to quantify species distributions?

- Presence-only data: locations where species occur
- Presence-absence data: data takes value 1 if a species is present at some location and/or time point and 0 if the species is absent
- Count data: a non-negative integer value indicating the number of individuals of a species at some location and/or time point

What data do we use to quantify species distributions?

- Presence-only data: locations where species occur
- **Presence-absence data: data takes value 1 if a species is present at some location and/or time point and 0 if the species is absent**
- Count data: a non-negative integer value indicating the number of individuals of a species at some location and/or time point

Presence-absence data

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

Presence-absence data

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

Assuming no false positives, if we detect the species, we know it exists at the site

Presence-absence data

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

Assuming no false positives, if we detect the species, we know it exists at the site

A 0 (or nondetection) could mean:

Presence-absence data

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

Assuming no false positives, if we detect the species, we know it exists at the site

A 0 (or nondetection) could mean:

1. The species does not exist at the site

Presence-absence data

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

Assuming no false positives, if we detect the species, we know it exists at the site

A 0 (or nondetection) could mean:

1. The species does not exist at the site
2. The species exists at the site, but we failed to detect it.

Occupancy modelling

- Developed to more accurately estimate species distributions when *imperfect detection* (false negatives) exists (MacKenzie et al. 2002; Tyre et al. 2003)

Occupancy modelling

- Developed to more accurately estimate species distributions when *imperfect detection* (false negatives) exists (MacKenzie et al. 2002; Tyre et al. 2003)
- **Fundamental concept**: obtain "repeated surveys" at a given site during some period of closure

Occupancy modelling

- Developed to more accurately estimate species distributions when *imperfect detection* (false negatives) exists (MacKenzie et al. 2002; Tyre et al. 2003)
- **Fundamental concept**: obtain "repeated surveys" at a given site during some period of closure
  - Key assumption: the species does not move in or out of the site during this time period

Occupancy modelling

- Developed to more accurately estimate species distributions when *imperfect detection* (false negatives) exists (MacKenzie et al. 2002; Tyre et al. 2003)
- **Fundamental concept**: obtain "repeated surveys" at a given site during some period of closure
  - Key assumption: the species does not move in or out of the site during this time period
- "Repeated surveys" usually come in the form of multiple visits to a site during some time period, but can also take different forms (e.g., multiple observers, spatial replicates)

Data for occupancy modelling

Detection-nondetection matrix (y)

k

j

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

$y_{j,k}$

- *J* sites with K_j replicate surveys at each site *j*
- Assume no false positives
- Any variation in the observed data values across surveys is assumed to arise from imperfect detection.

Occupancy model structure

- Two distinct sub-models
  1. Model occupancy probability as a function of site-level covariates

Occupancy model structure

- Two distinct sub-models
  1. Model occupancy probability as a function of site-level covariates
  2. 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

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

Occupancy (ecological) sub-model

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

z_j True occurrence of the species at site *j*

ψ_j Occurrence probability at site *j*

$X_{r,j}$ The *r*th covariate at site *j* (e.g., habitat variable)

Single-species occupancy model

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

Detection (observation) sub-model

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

$y_{j,k}$ Detection-nondetection data at site *j* during replicate *k*

$p_{j,k}$ Detection probability at site *j* during replicate *k*

$V_{r,j,k}$ Covariate affecting detection at site j during replicate k

Single-species occupancy model

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

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

- One logistic regression for occupancy probability
- One logistic regression for detection probability (conditional on occupancy)

Assumptions of the basic occupancy model

1. Closure assumption
2. No false positive errors
3. Independence of occurrence and detection
4. No unexplained heterogeneity in detection probability
5. Parametric assumptions (i.e., our model fits the data well)

Bayesian Basics

Why Bayesian for spatial occupancy models?

1. Interpretation

Why Bayesian for spatial occupancy models?

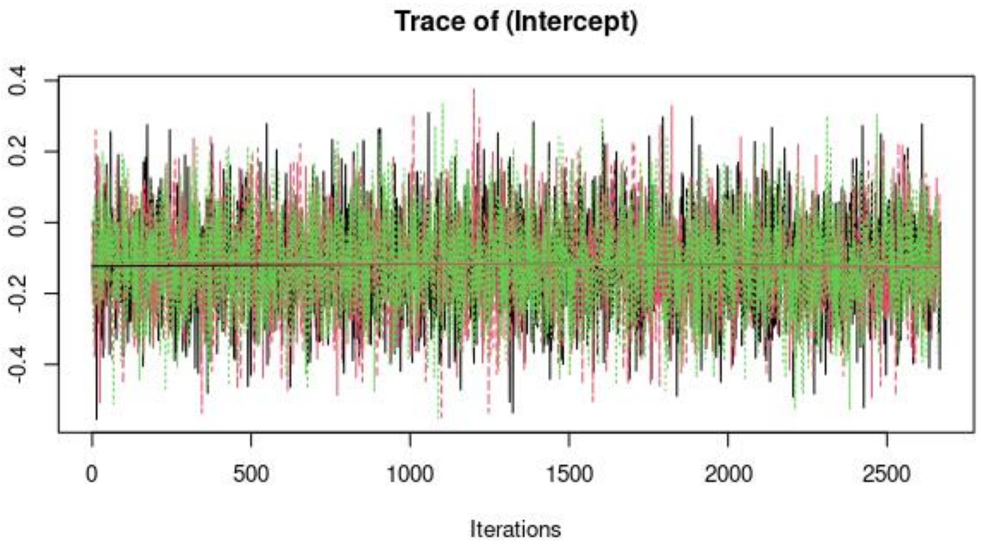
1. Interpretation
2. More flexible to accommodate spatial autocorrelation

Why Bayesian for spatial occupancy models?

1. Interpretation
2. More flexible to accommodate spatial autocorrelation
3. Easy to extend to multispecies frameworks/integrate multiple data sources

Bayesian basics: what to know to get started in spOccupancy

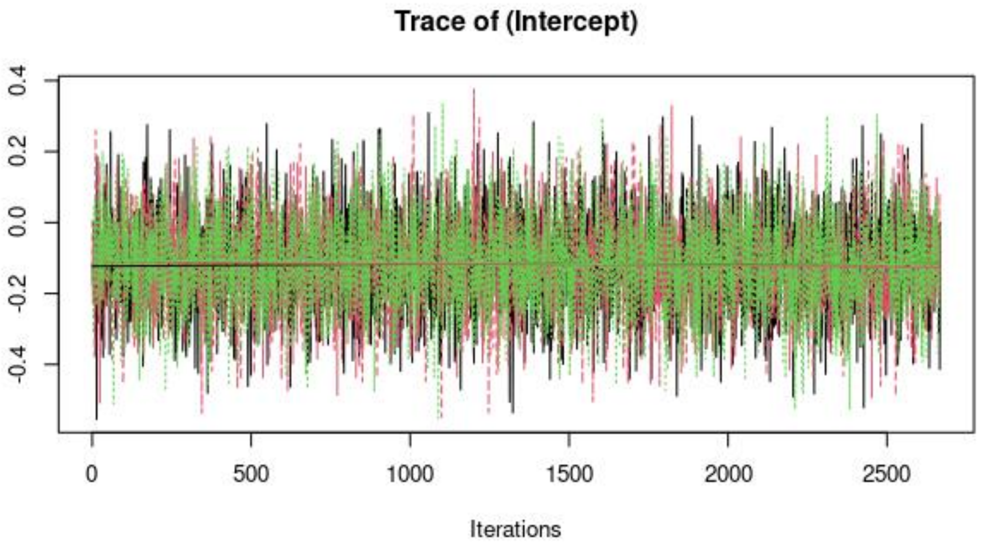
- Markov chain Monte Carlo (MCMC)
- MCMC chains eventually converge to a posterior distribution
  - Assess convergence by running multiple chains with different starting values



Bayesian basics: what to know to get started in spOccupancy

Monte

- Markov chain Monte Carlo (MCMC)
- MCMC chains eventually converge to a posterior distribution
  - Assess convergence by running multiple chains with different starting values



MCMC Step 1: Specify prior distributions

$$\beta \sim \text{Normal}(\mu_\beta, \sigma_\beta^2)$$
$$\alpha \sim \text{Normal}(\mu_\alpha, \sigma_\alpha^2)$$

MCMC Step 2: Set initial values

- Set different values for each chain
- spOccupancy will set initial values by default
- Can be important for more complicated models (e.g., spatially-varying coefficient models)

MCMC Step 3: Propose new value

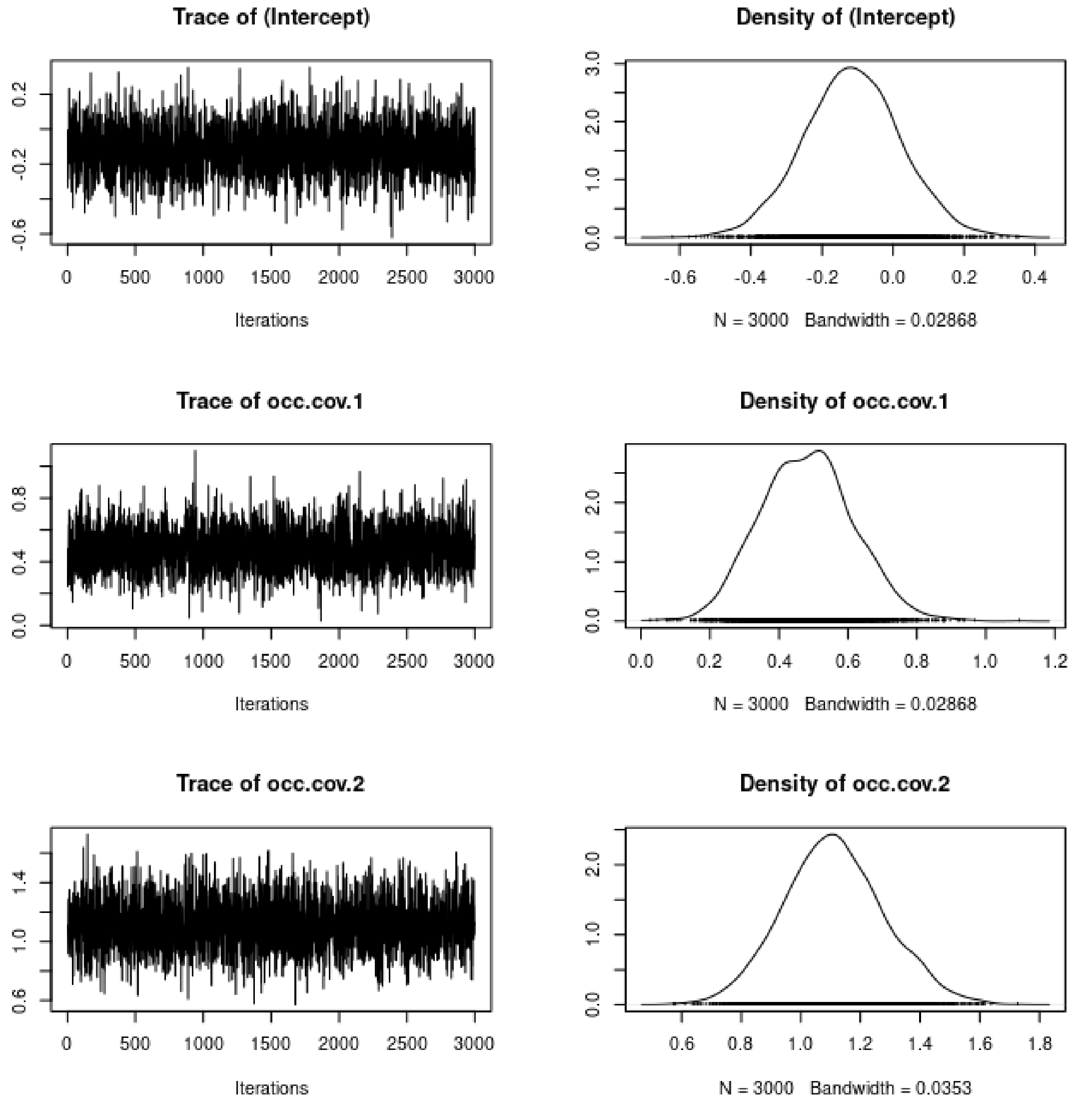
- Propose a new value for each parameter one at a time based on a statistical algorithm.
- For some parameters, we always accept the proposed value because our algorithm is efficient.
- For parameters with less efficient algorithms, we will accept the new value with some probability p.

MCMC Step 4: Repeat

- Repeat step 3 "many" times to generate a set of samples from the posterior distribution for each parameter.

MCMC Step 5: Summarize

- Point estimate: mean, median, mode
- Uncertainty: 95% credible interval (e.g., 2.5 and 97.5% quantiles of the samples)



What do you need to specify?

- Prior distribution (optional)
- Initial values (optional)
- Number of samples/iterations
- Burn-in: initial part of the MCMC chain that we throw away
- Thinning rate: how often do you want to save a sample?

spOccupancy

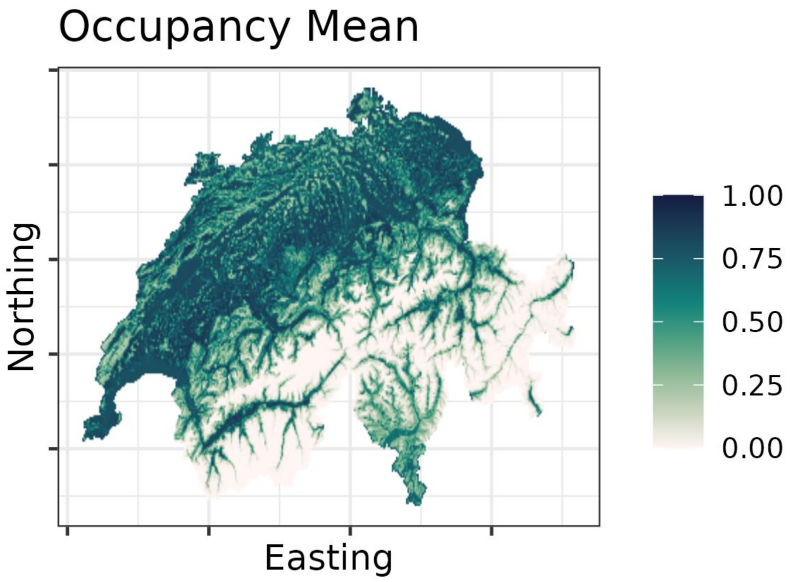
- Designed to fit Bayesian single-species and multi-species occupancy models
- Efficient options (NNGPs) to account for spatial autocorrelation
- Workflow completely in R (no Bayesian programming languages necessary)
- `PGOcc` -> single-species occupancy model
- `spPGOcc` -> spatial single-species occupancy model
- The "PG" stands for Pólya-Gamma (Polson et al. 2013)

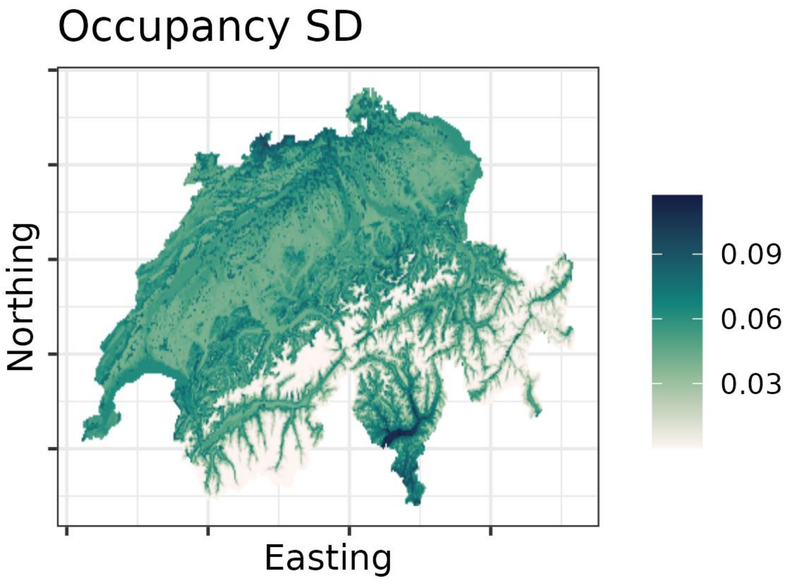
`spOccupancy` workflow

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

Exercise: European goldfinch distribution across Switzerland

- Data come from the Switzerland Breeding Bird Survey in 2014 (Swiss MHB)
- 266 survey locations distributed throughout Switzerland
- Objective: generate a species distribution map across the country





Exercise: Occupancy modeling of the European goldfinch

1-swiss-european-goldfinch.R