Introduction to Applied Bayesian Analysis in Wildlife Ecology

Jeffrey W. Doser

May 11, 2024

Bonus: occupancy models

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 modeling 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 modeling species distributions:
  - Imperfect detection
  - Spatial autocorrelation

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 modeling

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

Occupancy modeling

- 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 modeling

- 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 modeling

- 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 modeling

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

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)

Example: European goldfinch distribution across Switzerland

8-occupancy-models.R