Lecture 26 Occupancy (Presence /Absence) Estimation

Listing of Major Ecological Metrics

Single Season - Occupancy Metric

Multiple Seasons - Changes in Occupancy Metrics

Single Season Occupancy this lecture

Ecological Metrics: Different Choices for Different Situations

Individual Species Metrics

- Population Metrics (Earlier Lectures)
- Presence / Absence or Occupancy Metric (This Lecture)

Community Metrics (Recent Lectures)

- Species Richness
- Species Diversity
- Species Abundance Curves

One Common Theme in the Recent Research on all Metrics is the Need To Account For Detection Probability

Patch Occupancy Rate Metric: The Problem

- Conduct "presence-absence" (detection-non detection) surveys for a particular species of interest. Estimate what fraction of sites (or area) is occupied by a species (ψ) when the species is not always detected with certainty, even when present (i.e. $p < 1$)

- Naïve Occupancy rate estimates are biased low because some sites where the species was not detected are occupied.

- In other words apparent absence of a species from a site may just be a failure to detect the species. This could be be because individuals are hard to detect or because the species is very rare or both!!

- Also can we detect changes in the occupancy rate temporally or spatially.

Patch Occupancy Metric: Motivation

Some Reasons why the Information may be Needed

- Extensive Monitoring Programs
- Geographic Range Changes (climate change, habitat fragmentation, pollution)
- Meta Population Processes
- Habitat Selection
- Invasive Species Spread

Model Parameters

ψ_i -probability site *i* is occupied

p_{ij} -probability of detecting the species in site *i* at time *j*, given species is present

Note-Both may be functions of patch specific covariates

Patch Occupancy Metric: Key Design Issue is "Replication" to Estimate Detection Probability

- Replication is crucial if we are to separate occupancy from detection probability. There are several types of replication possible.

- Usual method is **temporal replication**: several repeat visits to sample units within a relatively short period of time (e.g., a breeding season)

- **Spatial replication**: randomly select a subsample of sites within each sample unit

- **Observer replication**: have several observers go to each site independently

Data Summary: Detection Histories

- A detection history (h_i) for each visited site or sample unit (i)
  - 1 denotes detection
  - 0 denotes nondetection
- Example detection history: h_i = 1 0 0 1
  - Denotes 4 visits to site
  - Detection at visits 1 and 4, non detection at 2 and 3

Data Summary: Detection Histories: 4 visits to S sites. S is known.

Site	Detection History
1	1101
2	1001
3	1100
4	0000- Note detections histories with all 0s allowed (Differs in capture-recapture)
.	
.	
S	1011

A Probabilistic Model: Very Similar to Capture-Recapture Models in Concepts Used

Sites that are occupied, For example

Pr(detection history 1001) = Pr (h_i = 1001) =

$$\psi_i \left[p_{i1} \left(1 - p_{i2}\right)\left(1 - p_{i3}\right) p_{i4} \right]$$

Model: Key Issue, Apparent vs. True Absence

Sites where the species was not detected at all. These sites may or may not be occupied

Pr(detection history 0000) = Pr (h_k = 0000) =

$$\psi_k \prod_{j=1}^{4} \left(1 - p_{kj}\right) + \left(1 - \psi_k\right)$$

First Term- Site is Occupied but Species Escapes Detection

Second Term- Site is Unoccupied

3 Visit Example: All Cell Types

If 3 visits then there are 8 capture histories with expected numbers

111 $N\psi\, p_1 p_2 p_3$

110 $N\psi\, p_1 p_2 (1-p_3)$

101 $N\psi\, p_1 (1-p_2) p_3$

011 $N\psi\, (1-p_1) p_2 p_3$

100 $N\psi\, p_1 (1-p_2)(1-p_3)$

010 $N\psi\, (1-p_1) p_2 (1-p_3)$

001 $N\psi\, (1-p_1)(1-p_2) p_3$

→ 000 $N[\{\psi\,(1-p_1)(1-p_2)(1-p_3)\}+(1-\psi)]$

We can estimate the parameters $\{\psi, p_1, p_2, p_3\}$ or submodels using ML methods. We could also make ψ and p functions of covariates

Model: The Likelihood Function

- The combination of these statements forms the model likelihood:

$$L(\underline{\psi}, \underline{p} \mid h_1, h_2, \ldots h_S) = \prod_{i=1}^{S} \Pr(h_i)$$

- Maximum likelihood estimates of parameters can be obtained
- Var-Cov matrix estimated using inverse of Fisher Information or parametric bootstrap
- However, parameters cannot be site-specific without additional information (covariates)

Model: Assumptions

- The detection process is independent at each site
- No heterogeneity that cannot be explained by covariates
- Sites are closed to changes in occupancy state between sampling occasions

Model: Covariates

- Site-specific: model ψ and/or p
  – e.g., habitat type, patch size, patch isolation
- Survey-specific: model p
  – e.g., local environmental conditions (water temperature)
- Use link function such as logistic. Here is an example for:

$$p_{ij} = \frac{e^{\beta_0 + \beta_1 x_i + \beta_2 y_{ij}}}{1 + e^{\beta_0 + \beta_1 x_i + \beta_2 y_{ij}}}$$

Model: 'Missing' Observations

- Implicit assumption that j^{th} surveys for all units are conducted at (approximately) the same time; possibly unlikely in practice
- Weather or breakdowns may result in some units not being surveyed
- Equal sampling effort may not be possible with available resources

‘Missing’ Observations

	Day				
Unit	1	2	3	4	5
1	1	0	1	-	0
2	-	0	-	1	1

$$\Pr\left(\mathbf{H}_1 = 101-0\right) = \psi p_1\left(1-p_2\right)p_3\left(1-p_5\right)$$

$$\Pr\left(\mathbf{H}_2 = -0-11\right) = \psi\left(1-p_2\right)p_4 p_5$$

Model: Software

- Windows-based software:
  - Program PRESENCE (J. Hines &D. MacKenzie) is available at the Patuxent Software Site. We will demonstrate in a later lecture.
  - Program MARK (G. White).
- Fit both predefined and custom models, with or without covariates
- Provide maximum likelihood estimates of parameters and associated standard errors
- Assess model fit

Simple Example: Anurans at Maryland Wetlands (Droege and Lachman)

- Frogwatch USA (NWF/USGS)
- Volunteers surveyed sites for 3-minute periods after sundown on multiple nights
- S = 29 wetland sites; Piedmont and Coastal plain
- 27 Feb. – 30 May, 2000
- Covariates:
  - Sites: habitat ([pond, lake] or [swamp, marsh, wet meadow])
  - Sampling occasion: air temperature

Example: Anurans at Maryland Wetlands (Droege and Lachman)

- American toad (*Bufo americanus*)
  - Detections at 10 of 29 sites (Naïve = 10/29=0.34)

Example: Anurans at Maryland Wetlands *(B. americanus)*

Model	ΔAIC	$\hat{\psi}$	$\hat{SE}(\hat{\psi})$
$\psi(\text{hab})p(\text{tmp})$	0.00	0.50	0.13
$\psi(.)p(\text{tmp})$	0.42	0.49	0.14
$\psi(\text{hab})p(.)$	0.49	0.49	0.12
$\psi(.)p(.)$	0.70	0.49	0.13

Naive $\hat{\psi} = 0.34$

Example: Anurans at Maryland Wetlands (*B. americanus*)

- Notice that the naïve occupancy estimates have a serious negative bias (0.34 vs. 0.5).
- Notice that choice of model has little effect on the estimate or its SE.
- Notice that S=29 is a very small no. of sites to monitor!! Thus the SEs are quite large under all the different models. The largest Relative SE is about 28 % (0.14/0.49). Also this is why the AIC is having trouble distinguishing between the models.
- Clearly this estimate will be of most use an ecologist if compared to estimates in different years in a monitoring program to detect trends in occupancy.

Model assumptions

- Closure
- Surveys are independent
- No unmodeled heterogeneity

Closure: What if Occupancy Changes during the Survey?

- Is a ‘season’ defined appropriately? The max length of time that is reasonable will depend on the species.

Lack of Independence

- Surveys are not independent if the outcome of survey A is dependent upon the outcome of survey B.
- Usually, parameter estimates may be OK, but standard errors too small

Unmodeled Heterogeneity

In occupancy probabilities

- parameter estimates should still be valid as average values across the sites surveyed

In detection probabilities

- occupancy will be underestimated (similar to capture-recapture studies)
- covariates may account for some sources of variation

Abundance induced Heterogeneity

- Differences in the local abundance of the species between sites may induce heterogeneity in detection probability
- Royle and Nichols (2003) suggested an extension of the above method to accommodate this (location *i*, time *j*)

$$p_{ij}^{(N)} = 1 - \left(1 - p_j\right)^{N_i}$$

Design Issues for Occupancy Studies

- What is a ‘season’?
- How to define a ‘sampling unit’?
- Selecting sampling units
- Repeat surveys
- Avoiding heterogeneity
- More units vs. more surveys?

What is a ‘Season’?

- A season is a period of time during which it is reasonable to assume occupancy is static or changes occur completely at random

- Depends very much on the target species and study objective

How to define a ‘sampling unit’?

- Should be assessed on a case-by-case basis
- Large enough to have a reasonable probability of occupancy, but not so large that any measure may be meaningless
- Size matters!

How to define a ‘sampling unit’?

- Is there a natural definition?
- At what scale do you want to measure occupancy?
- Is the species territorial?
- What density does it occur at?
- What is the size of it’s home range?

Selecting Sampling Units

- How units are selected determines how results can be generalized
- Each sampling unit within the population should have a non-zero probability of being selected
- If units are selected such that occupancy is different than for the population of interest, estimates may be biased.
  - e.g., surveying only at historic sites

Repeat Surveys

Repeat surveys do not (necessarily) imply repeat visits

- Discrete visits
- Multiple surveys within single visit
  - Single observer, conducting multiple surveys
  - Multiple observers each conducting a single survey
  - Multiple survey plots within a larger sampling unit

Allocation of Effort

- For a ‘standard design’ there is an optimal number of repeat surveys per unit.
- The optimal number depends upon values for ψ and p.
- Does not depend upon number of units or total number of surveys.
- Reasonably robust to effect of cost.

Number of Surveys

	ψ								
p	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
0.1	14	15	16	17	18	20	23	26	34
0.2	7	7	8	8	9	10	11	13	16
0.3	5	5	5	5	6	6	7	8	10
0.4	3	4	4	4	4	5	5	6	7
0.5	3	3	3	3	3	3	4	4	5
0.6	2	2	2	2	3	3	3	3	4
0.7	2	2	2	2	2	2	2	3	3
0.8	2	2	2	2	2	2	2	2	2
0.9	2	2	2	2	2	2	2	2	2

How many Units?

- Once the number of repeat surveys has been determined, how many units to survey can be determined from the variance equation

$$Var\left(\hat{\psi}\right)=\frac{\psi}{U}\left[\left(1-\psi\right)+\frac{\left(1-p^{*}\right)}{p^{*}-Kp\left(1-p\right)^{K-1}}\right]$$

$$p^{*}=1-\left(1-p\right)^{K}$$

How many Units?

- Example, if $\psi \approx 0.7$ and $p \approx 0.4$, should use 5 surveys per unit; $p^* = 0.92$
- To achieve a SE of 0.04 use S=183 based on Equation 6.1 in the reference book and on the previous equation.

General Recommendations on Allocating Effort

- When detection probability is >0.5, at least 3 surveys per unit
- More surveys will be required when *p* is lower
- For rare species, survey more units less intensively
- Increasing spatial replication with insufficient repeat surveys may not be worthwhile

General recommendations on allocating Sampling Effort

Example: if $\psi \approx 0.4$ and $p \approx 0.3$

Surveying 200 units twice gives $SE(\hat{\psi}) = 0.11$

Surveying 80 units 5 times gives $SE(\hat{\psi}) = 0.07$

- a decrease of 36%

With only 2 surveys per unit, would require 500 units to achieve same level of precision, or increase total effort 250%

Non-Standard Designs?

- Repeatedly surveying a subset of units and elsewhere only once *does not* generally provide a more efficient design.
- Surveying a unit repeatedly until first detection (up to a maximum) may provide a more efficient design, but may be less robust. (Note the ‘optimal’ maximum number is higher than values given in the previous table)
- The standard design where each site visited the same no of times is probably best.

Final Comments on Design

- Designing studies tends to be an iterative affair
- Simulation and pilot studies can provide useful information on how designs and field methods are likely to perform
- Program GENPRES (Hines) is available at the Patuxent Software Site.

Examples: Pronghorn Antelope Data

- Complete analysis given on P 113 in Book. Habitat covariates.
- One important point is that analyses ignoring detection probability (simple logistic regression on the apparent presence absence data) can be very misleading if the detection probability.
- It is much better to use the occupancy approach which can be viewed as a generalisation of logistic regression and allows for uncertain detection.

Other Examples: Giant Weta Study

- An ancient giant insect (Order Orthoptera) now endangered in NZ due to introduction of small rodents. P116 in Book.
- Therefore the need to look at occupancy within a reserve and how it is affected by habitat (browse vs non browse by cattle and goats)
- 72 3m radius plots searched between 3 and 5 times in March 2004

Other Examples: Giant Weta Analysis

Occupancy covariate - browse or no browse

Detection probability covariates - day and observer

Best Model using AIC is $\psi(\text{Browse}),\, p(\text{day}+\text{observer})$

Other Examples: Giant Weta Analysis

Occupancy Estimation

Naive $\hat{\psi} = 0.49$ (SE = 0.06)

$\hat{\psi}(\text{Browse}) = 0.77$

$\hat{\psi}(\text{No Browse}) = 0.50$

$\hat{\psi} = 0.63$ (SE = 0.08)

Detection Probability Estimation

$\hat{p}$ varied widely from 0.10 - 0.69.

That is why the occupancy estimate of 0.63 is so much higher than the naive occupancy estimate of 0.49.

Occupancy Estimation: Reference

MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. L., and Hines, J. E. (2005).

Occupancy Estimation and Modeling : Inferring Patterns and Dynamics of Species Occurrence.

Elsevier, San Diego, USA.

Make Sure You all Run Out and Buy It!

Next Lecture: Multiple Seasons: Changes in Occupancy

- There are multiple visits within each of several seasons (Fig 7.1from Mackenzie et al. 2005)
- Within a season we have closure while between seasons there can be local extinction (ε) and also local recolonisation (γ) (Fig 7.1 and 7.2).
- This is similar in concept to the robust design used for capture-recapture(Ch 19 Text) and for looking at changes in communities(Ch 20 text)).

Generalised Detection Histories

- For illustration of the approach attached is an example derivation of the detection history cell probability structure of a site over 3 seasons with 3 visits per season

  110 000 010
- Parameters are initial occupany, extinction, recolonisation, and detection probabilities
- As for the single season models, likelihoods can be constructed for submodels, best models can be found using AIC and then estimates computed for the best model

House Finch Expansion

- 1942 a small release of house finches was made on Long Island NY. Impressive western range expansion in the eastern US since then.
- On P 201 in Mackenzie et al. (2005) they give a detailed example of occupancy change, and colonisation over year since invasion and distance band from the release point. Details not discussed here.