Occupancy Modeling

Reading:

MacKenzie et al. 2002

Announcements

- Monday, April 1 – No Class
- Wed, April 3 - In-class exam 2
- Tues, April 9 – Take-home exam 2 due (will have 10-14 days to complete.

Occupancy

- Occupancy (ψ)
  - Measuring the presence or absence of a species in a location
  - (Alternative to abundance)
- Uses for occupancy data/studies? Examples?

Occupancy

- Occupancy (ψ)
  - Measuring the presence or absence of a species in a location
  - Given as a probability or frequency
  - (Alternative to abundance)
- Uses for occupancy data/studies? Examples?
  - Studies of species distributions
    - What factors affect presence?
    - Habitat modeling
  - Metapopulation dynamics
    - where site (or patch) occupancy is related to site (or patch) characteristics
    - Can look at extinction and colonization probabilities

Occupancy

- How might someone estimate occupancy (ψ) if we had data from a survey at *n* sites?

$$\psi = \frac{n_{occupied}}{n_{total}}$$

Problem with this?

Problem?

- Was species *present and not detected* or was it absent?
- The measure of occupancy is **confounded** with the detectability of the species
  - **Confounded**: an inability to separate multiple factors potentially contributing to an observed pattern.
- Using presence/absence data alone will UNDERESTIMATE or OVERESTIMATE true occupancy.

Problem?

- Was species *present and not detected* or was it absent?
- The measure of occupancy is **confounded** with the detectability of the species
    - **Confounded**: an inability to separate multiple factors potentially contributing to an observed pattern.
- Using presence/absence data alone will UNDERESTIMATE or OVERESTIMATE true occupancy.

- **Detectability**
  - refers to the reality that it is very common for animals and even entire species to be missed and go undetected.
  - Expressed as a probability of detection

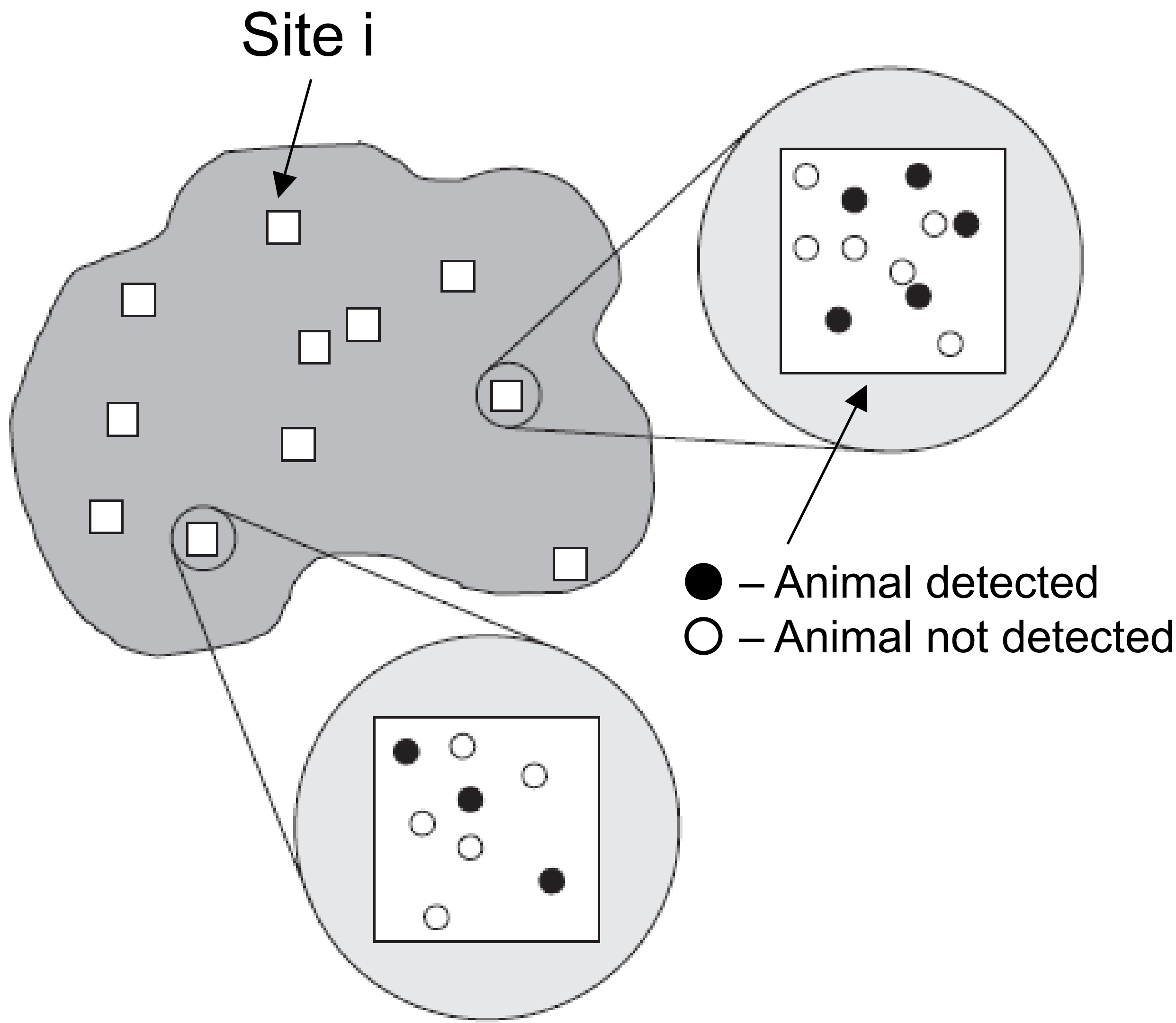


FIGURE 1.1 Illustration of the two critical aspects of sampling animal populations, spatial variation and detectability. The shaded region indicates the area or population of interest, with the small squares representing the locations selected for sampling. Within each sampling location, animals will be detected (filled circles) or undetected (hollow circles) during a survey or count.

Occupancy models

- **Models that deal with the problem of imperfect detectability**
- **Need info from repeated observations at each site to estimate detectability!**
  - Can be multiple visits, multiple observers, single observer & multiple passes, etc.
- Detectability can vary with <u>site</u> characteristics (e.g., habitat variables) or <u>survey</u> characteristics (e.g., weather)
  - Occupancy relates only to site characteristics

Effect of imperfect detection

Consensus models (assume perfect detection)

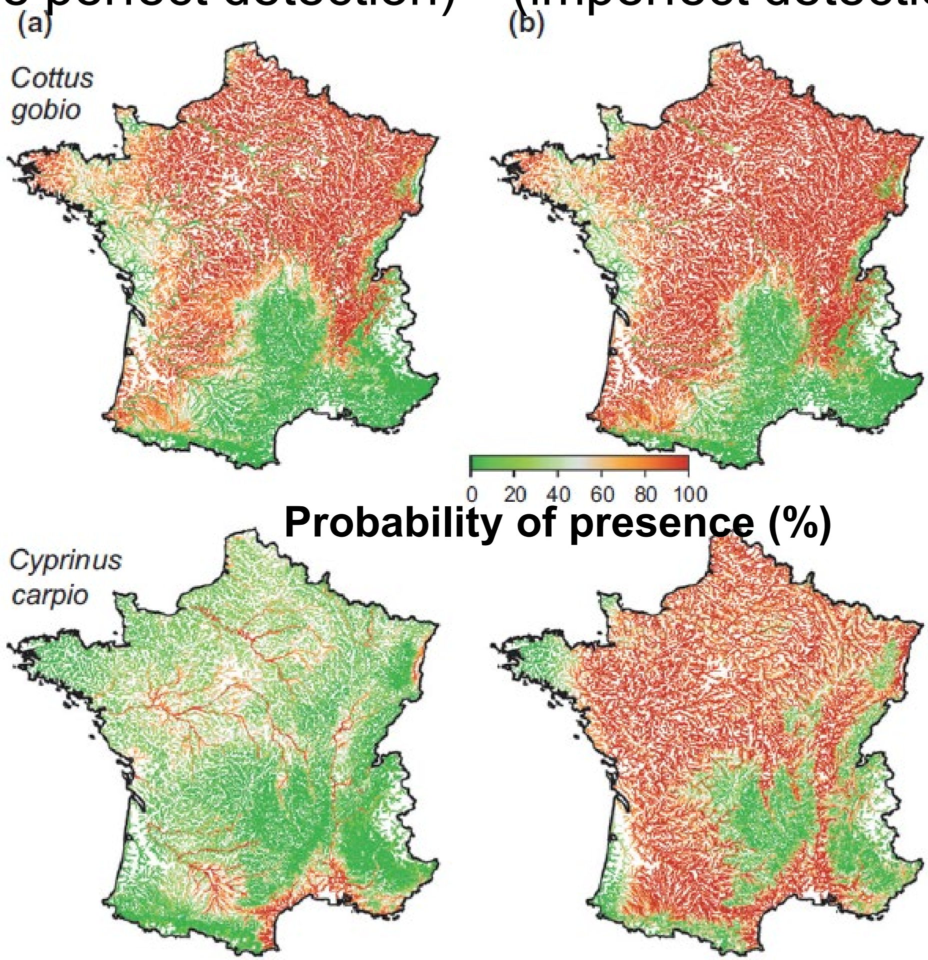
Occupancy models (imperfect detection)

European bullhead

High detectability

Common carp

Low detectability



What patterns do you see?

Comte and Grenouillet 2013

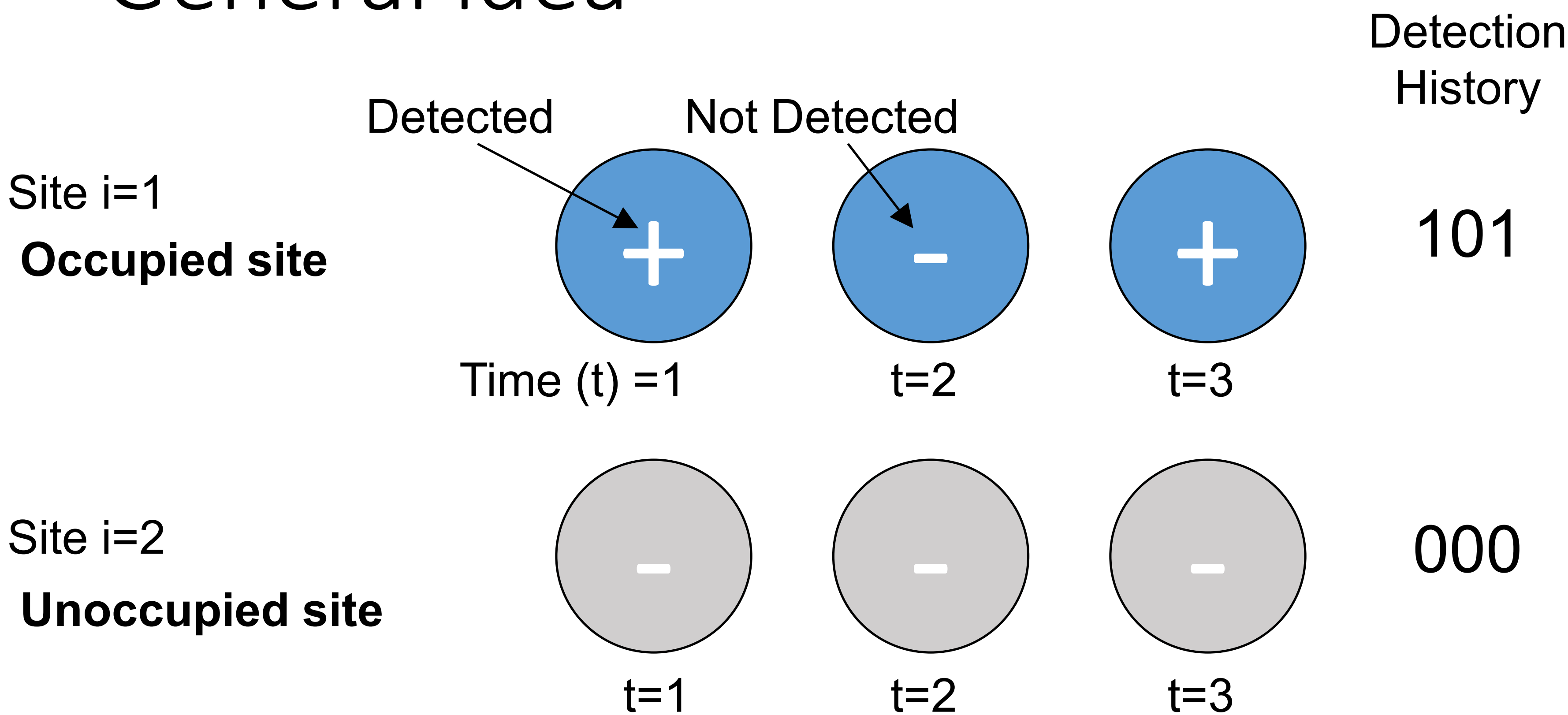
Next installment of “Google This”!

- “occupancy modeling funny”

Data needs

- Detection history
  - record of pres/abs data for each site (multiple occasions)
- Other variables of interest
  - Environmental variables pertaining to site
  - Variables pertaining to detection during repeat sampling events or surveys (e.g., weather)

General idea



Main parameters

- **ψ – probability of occupancy** = probability of the species being present at site i
- **p – conditional detection probability** = probability of detecting the species, given that it was present
  - This takes advantage of the multiples detections (e.g., over time t)

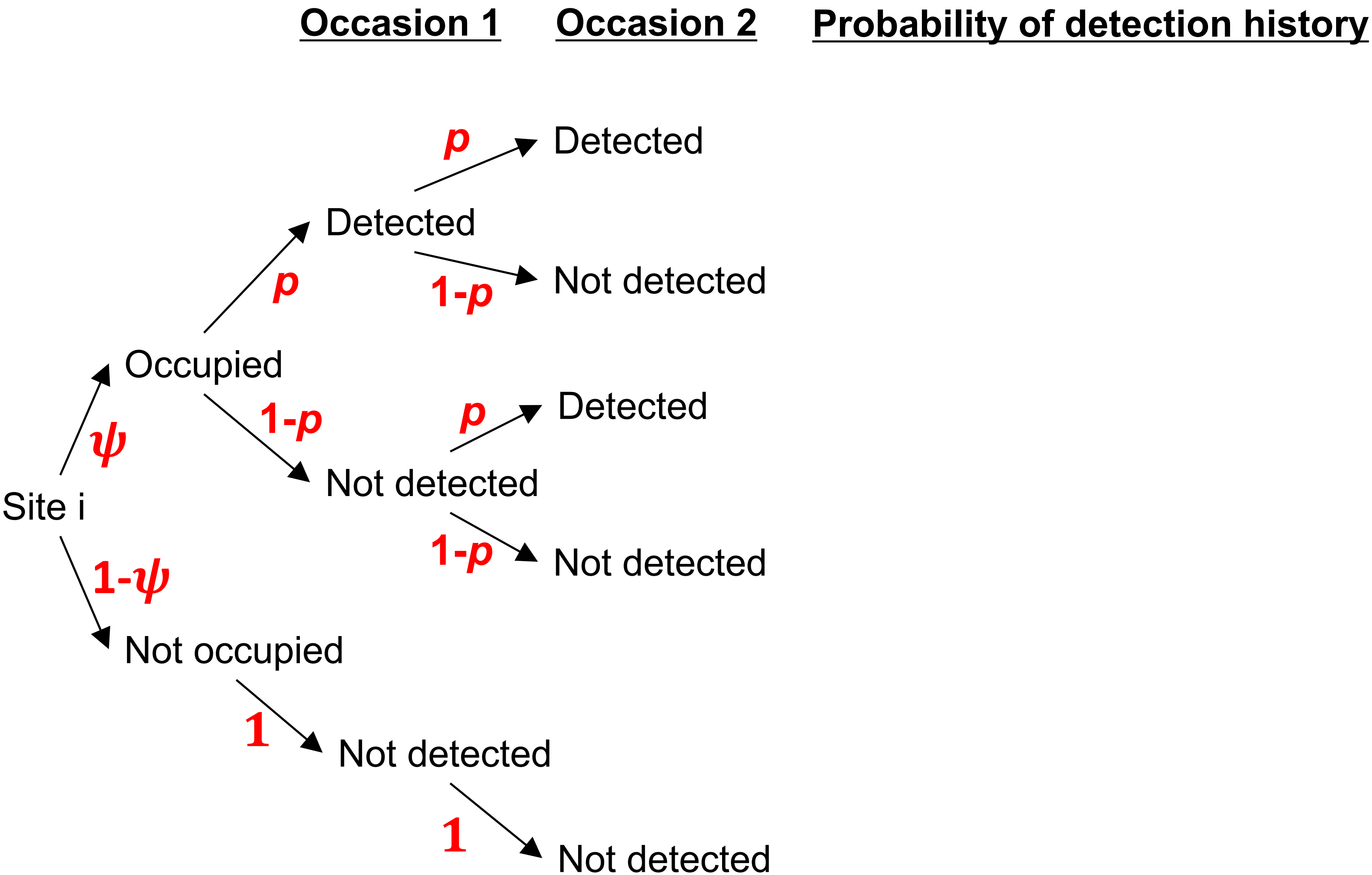
Single seas. occupancy (2 sampling occasions)

Occasion 1	Occasion 2	Probability of detection history

Site i

Modified from D. Barton

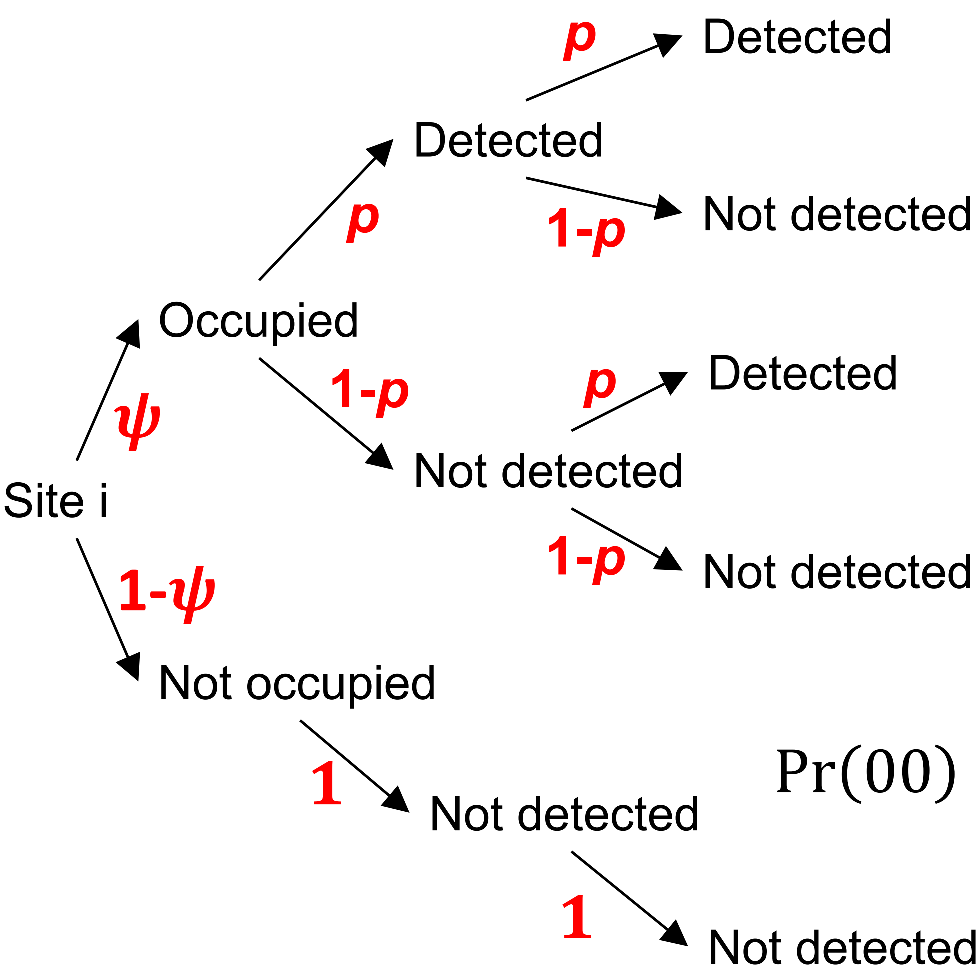
Single seas. occupancy (2 sampling occasions)



Modified from D. Barton

Single seas. occupancy (2 sampling occasions)

Occasion 1 **Occasion 2** **Probability of detection history**



$$\Pr(11) = \psi[p_1 p_2]$$

$$\Pr(10) = \psi[p_1(1-p_2)]$$

$$\Pr(01) = \psi[(1-p_1)p_2]$$

$$\cancel{\Pr(00)} = \psi[(1-p_1)(1-p_2)]$$

"false absence"

$$\Pr(00) = \psi[(1-p_1)(1-p_2)] + (1-\psi)$$

$$\cancel{\Pr(00)} = (1-\psi)$$

"true absence"

Modified from D. Barton

Draw hypothetical example…

Suppose you sample 5 sites to detect the endangered tidewater goby,
And you have 2 survey occasions per site:

Likelihood and log likelihood

Probability of detection history

n_{01} = number of sites with detection history of 01

$$\Pr(11) = \psi[p_1 p_2]$$

$$\Pr(10) = \psi[p_1(1 - p_2)]$$

$$\Pr(01) = \psi[(1 - p_1)p_2]$$

$$\Pr(00) = \psi[(1 - p_1)(1 - p_2)] + (1 - \psi)$$

- **Likelihood** of the dataset, given N total sites ($N=n_{11}+n_{10}+n_{01}+n_{00}$), is:

$$L(\psi, p|N) = \Pr(11)^{n_{11}} \Pr(10)^{n_{10}} \Pr(01)^{n_{01}} \Pr(00)^{n_{00}}$$

- And the **log likelihood** is:

$$\mathbf{log}\, \boldsymbol{L}(\boldsymbol{\psi}, \boldsymbol{p}|\boldsymbol{N}) = n_{11}\log(\psi[p_1 p_2]) + n_{10}\log(\psi[p_1(1 - p_2)]) + n_{01}\log(\psi[(1 - p_1)p_2]) + n_{00}\log[\psi[(1 - p_1)(1 - p_2)] + (1 - \psi)]$$

Modified from D. Barton

How to add covariate effects?

- Model ψ or p as a function of covariates (i.e., logistic regression)

Generic parameter (representing either ψ or p)

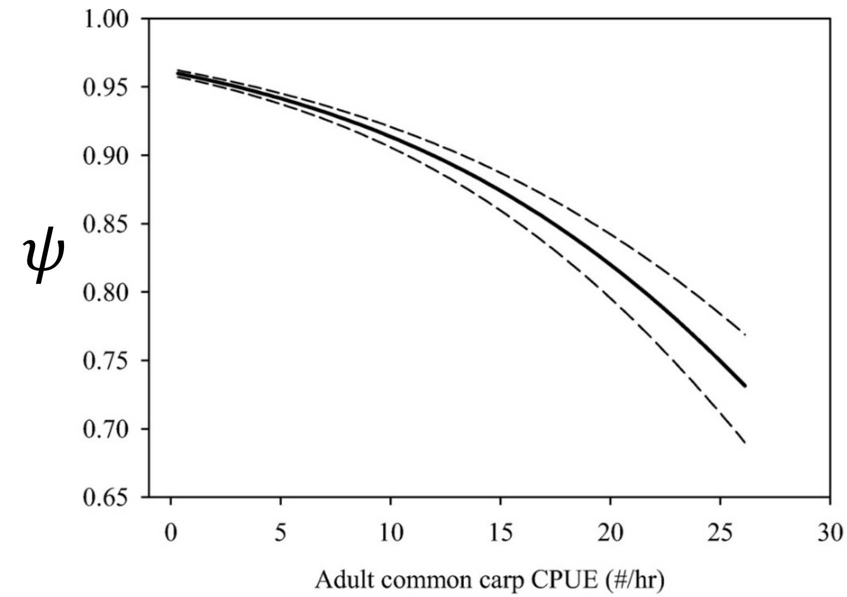
$$\theta = \frac{e^{\boldsymbol{X\beta}}}{1 + e^{\boldsymbol{X\beta}}} = \frac{e^{\beta_0 + \beta_1 X_1 \dots}}{1 + e^{\beta_0 + \beta_1 X_1 \dots}}$$

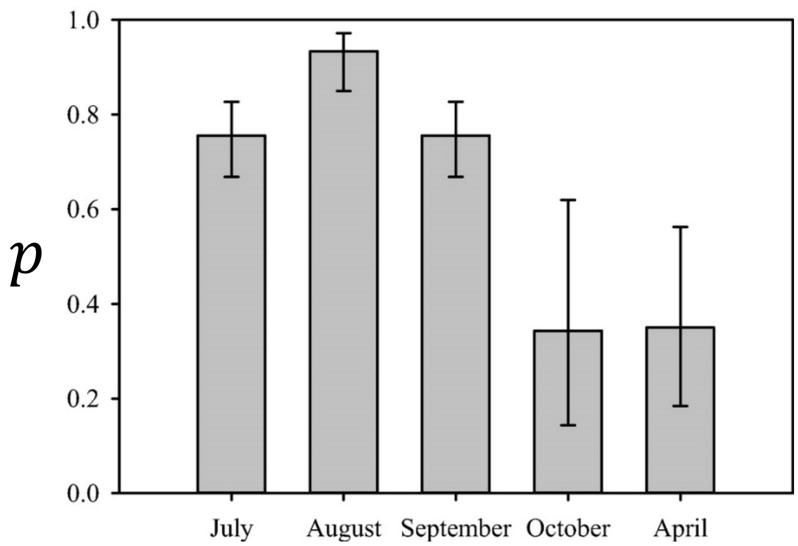
- Examples of covariates?
- Covariates can be for sites or for sampling occasions (aka surveys)

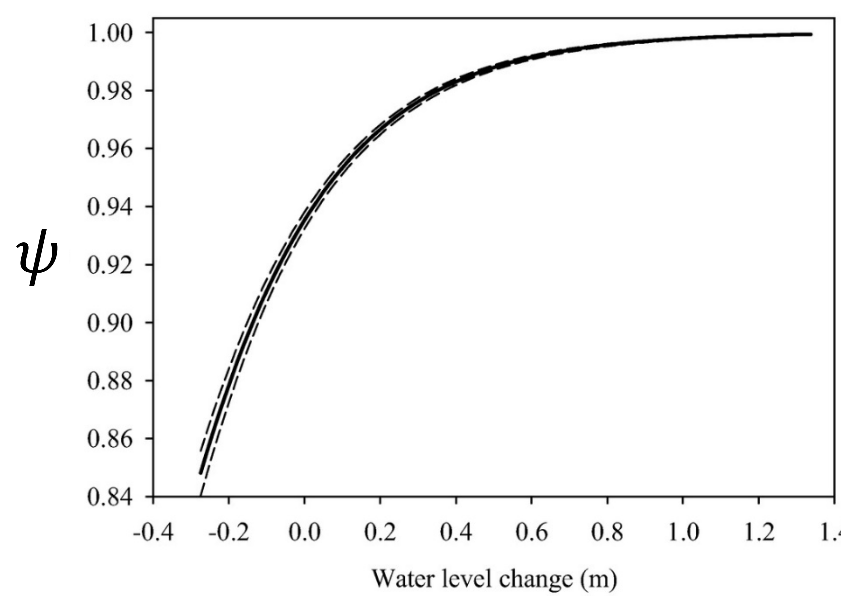
MacKenzie et al. 2002

Visualizing covariate effects

- Plot predicted psi and p as a function of your covariates (typically with 95% CI)
  - Can be continuous or categorical variables
- E.g.: Study on age-0 common carp occupancy in South Dakota







Weber and Brown 2019 (CJFAS)

How to add covariate effects?

- To get the average occupancy across sites (when modeled with covariates):
  - Need to average occupancy across covariate levels:

$$\overline{\hat{\psi}} = \frac{\sum_{i=1}^{N} \hat{\psi}_i}{N}$$

MacKenzie et al. 2002

Missing observations

- The model can handle missing observations
  - E.g., due to logistical issues or sampling design

$$\Pr(10\_11) = \psi p_1 (1 - p_2) p_4 p_5$$

- If you have some sites with only 1 sampling occasion:
  - Won't inform detection probability estimate,
  - but can apply estimates of p (and perhaps covariate effects) to those sites

Standard error of parameters

- Recommend bootstrapping
  - Resample N sites (with replacement) from the N sites
  - Refit model many times
  - Look at distribution of values
- Asymptotic standard error not recommended
  - ASE based on second order partial derivative of model likelihood

MacKenzie et al. 2002

Assumptions

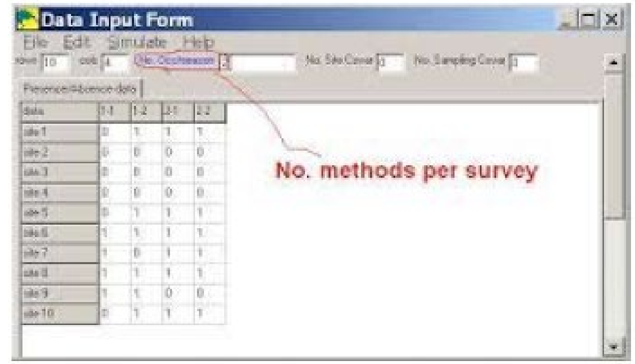
- Occupancy state is “closed” (no change in occupancy)
- Sites are independent
- No unexplained heterogeneity in occupancy
- No unexplained heterogeneity in detectability

How to meet assumptions?

- Occupancy state is “closed” (no change in occupancy)
  - Use life history or existing data to inform sampling
- Sites are independent
  - Select sites using probability based sampling scheme (e.g., simple-random; stratified-random)
- No unexplained heterogeneity in occupancy
  - Collect info about important variables
- No unexplained heterogeneity in detectability
  - Collect info about important variables

Occupancy model software

Program PRESENCE



Program MARK

R package Unmarked

Example

Spring peeper

American toad

- Covariates
  - Temperature
  - Habitat type
- 29 Sites
  - visited 2-66 times; mean ~9

MacKenzie et al. 2002

Example

What do the results of this table mean? (Recall "Lab07 – AIC background.pdf")

- Can compare evidence for different hypotheses (ie models)!

TABLE 1. Relative difference in AIC (ΔAIC), AIC model weights (w_i), overall estimate of the fraction of sites occupied by each species ($\hat{\bar{\psi}}$), and associated standard error (SE($\hat{\bar{\psi}}$)).

Model, by species	ΔAIC	w_i	$\hat{\bar{\psi}}$	SE($\hat{\bar{\psi}}$)
American toad				
ψ(Habitat) p(Temperature)	0.00	0.36	0.50	0.13
$\psi(\cdot)$ p(Temperature)	0.42	0.24	0.49	0.14
ψ(Habitat) $p(\cdot)$	0.49	0.22	0.49	0.12
$\psi(\cdot)$ $p(\cdot)$	0.70	0.18	0.49	0.13
Spring peeper				
ψ(Habitat) p(Temperature)	0.00	0.85	0.84	0.07
$\psi(\cdot)$ p(Temperature)	1.72	0.15	0.85	0.07
ψ(Habitat) $p(\cdot)$	40.49	0.00	0.84	0.07
$\psi(\cdot)$ $p(\cdot)$	42.18	0.00	0.85	0.07

MacKenzie et al. 2002

Simulation results

Take-homes

- Better performance with greater sampling occasions
- Lower precision with lower sampling occasions
- Poorer performance at lower p with T=2

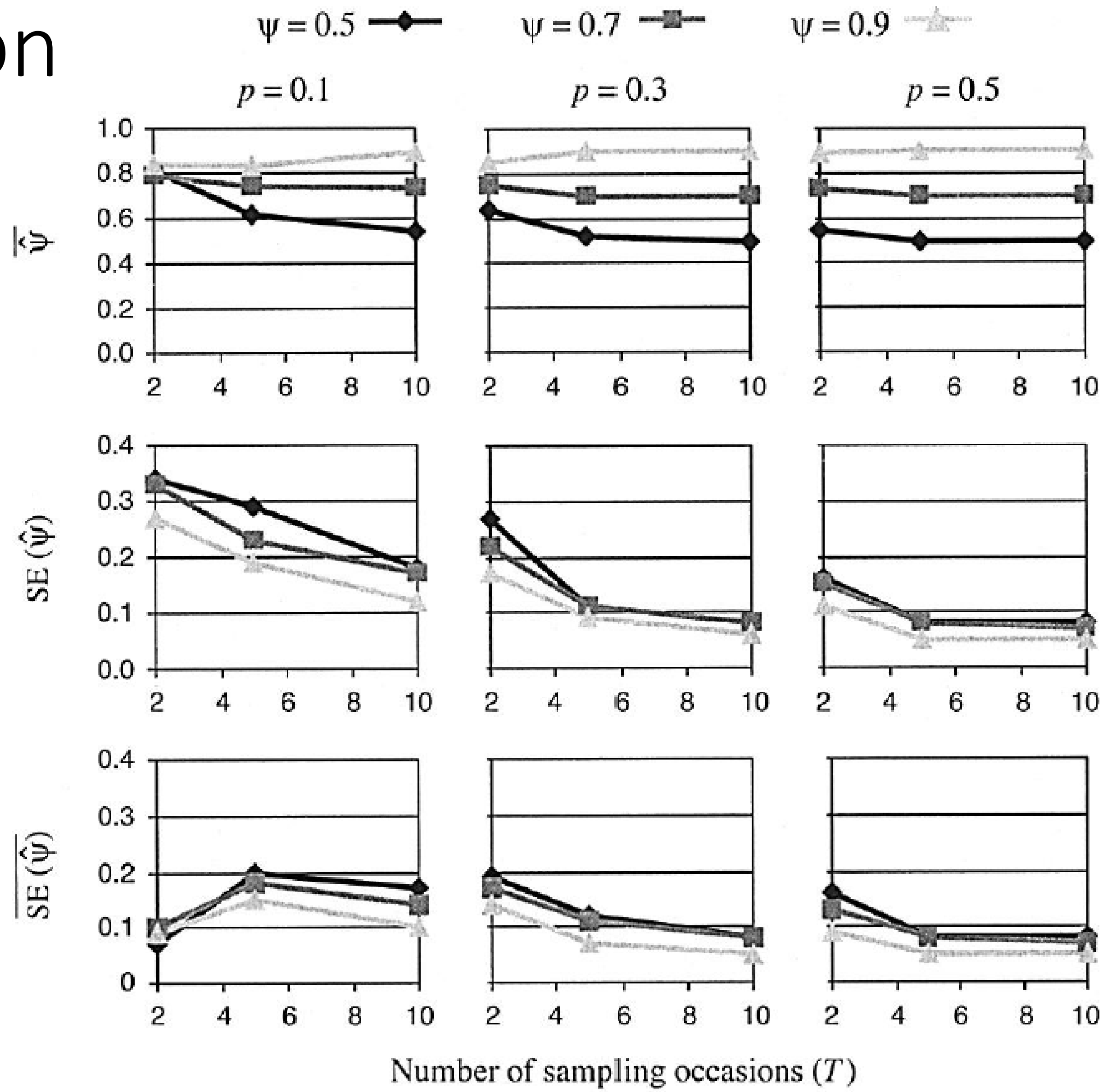


FIG. 1. Results of the 500 simulated sets of data for $N = 40$, with no missing values. Indicated are the average value of $\hat{\psi}$, $\overline{\hat{\psi}}$; the replication-based estimate of the true standard error of $\hat{\psi}$, $\mathrm{SE}(\hat{\psi})$; and the average estimate of the standard error obtained from 200 nonparametric bootstrap samples, $\overline{\mathrm{SE}}(\hat{\psi})$, for various levels of T, p, and ψ.

Considerations

- Recommended number of site visits?
  - More site visits = greater precision
  - If visiting only twice: best if occupancy is >0.7 and detection probability > 0.3
- Be skeptical of values if:
  - ψ close to 1, and p is low (<0.15)

MacKenzie et al. 2002

Extensions

- Having detection probability depend on abundance
  - Royle and Nichols 2003
- Multiple seasons
  - MacKenzie et al. 2003
- Multiple species
  - MacKenzie et al. 2004
- Multi-scale occupancy using multiple detection methods
  - Nichols et al. 2008
- …

MacKenzie et al. 2006 (book)

Ecology, 83(8), 2002, pp. 2248–2255

ESTIMATING SITE OCCUPANCY RATES WHEN DETECTION PROBABILITIES ARE LESS THAN ONE

DARRYL I. MACKENZIE,[1,5] JAMES D. NICHOLS,[2] GIDEON B. LACHMAN,[2,6] SAM DROEGE,[2] J. ANDREW ROYLE,[3] AND CATHERINE A. LANGTIMM[4]

Ecology, 84(8), 2003, pp. 2200–2207

ESTIMATING SITE OCCUPANCY, COLONIZATION, AND LOCAL EXTINCTION WHEN A SPECIES IS DETECTED IMPERFECTLY

DARRYL I. MACKENZIE,[1,5] JAMES D. NICHOLS,[2] JAMES E. HINES,[2] MELINDA G. KNUTSON,[3] AND ALAN B. FRANKLIN[4]

Journal of Animal Ecology 2004
73, 546–555

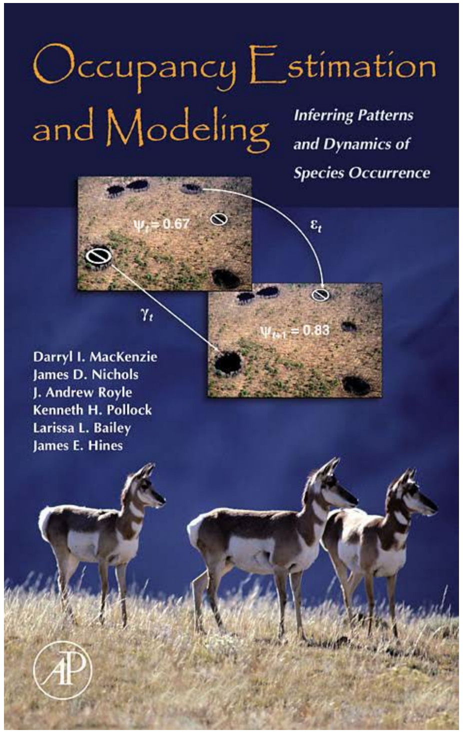
Investigating species co-occurrence patterns when species are detected imperfectly

DARRYL I. MACKENZIE*, LARISSA L. BAILEY† and JAMES. D. NICHOLS‡

Journal of Applied Ecology 2008, **45**, 1321–1329 doi: 10.1111/j.1365-2664.2008.01509.x

Multi-scale occupancy estimation and modelling using multiple detection methods

James D. Nichols[1]*, Larissa L. Bailey[1], Allan F. O'Connell Jr.[2], Neil W. Talancy[3], Evan H. Campbell Grant[1], Andrew T. Gilbert[4], Elizabeth M. Annand[5], Thomas P. Husband[3] and James E. Hines[1]



- MacKenzie et al. 2006 (book)
- Donovan, T. M. and J. Hines. 2007. Exercises in occupancy modeling and estimation. www.uvm.edu/rsenr/vtcfwru/spreadsheets/occupancy/occupancy.htm

E.g., Multi-season occupancy model

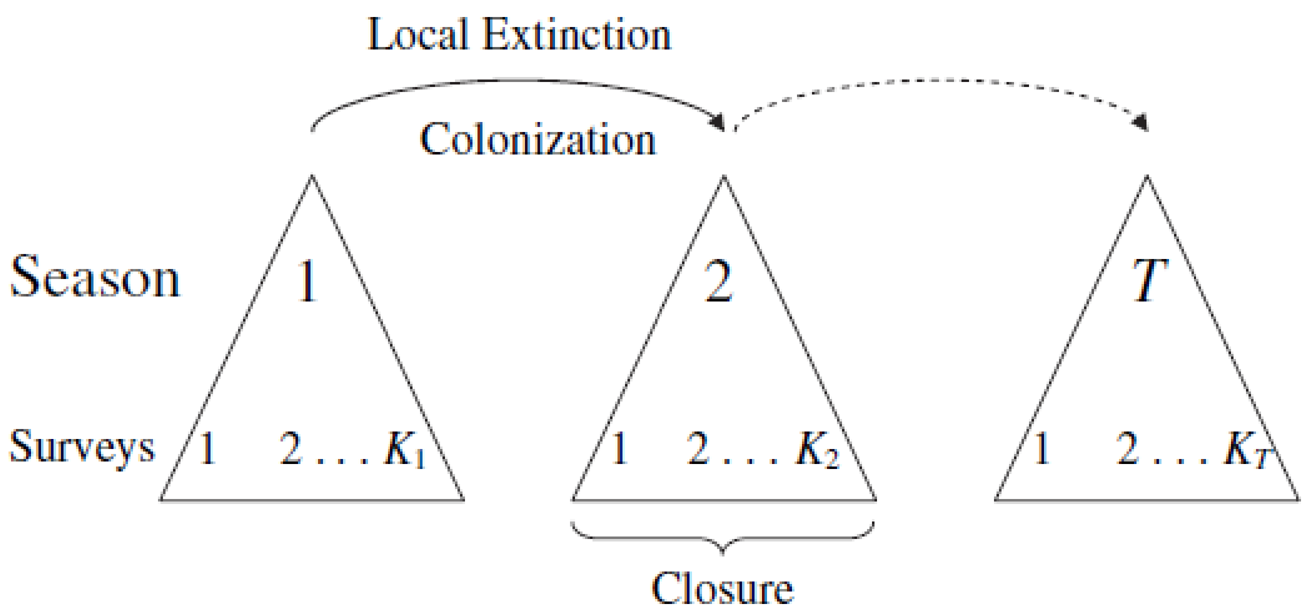


FIGURE 7.1 Graphical representation of the sampling situation for a multi-season occupancy study. Each triangle represents a season (t), with multiple (K_t) surveys within seasons. Sites are closed to changes in occupancy within seasons, but changes may occur between seasons through the processes of colonization and local extinction.

MacKenzie et al. 2006 (book)

Method is akin to "Pollock's robust design" for mark-recapture studies

E.g., Multi-season occupancy model

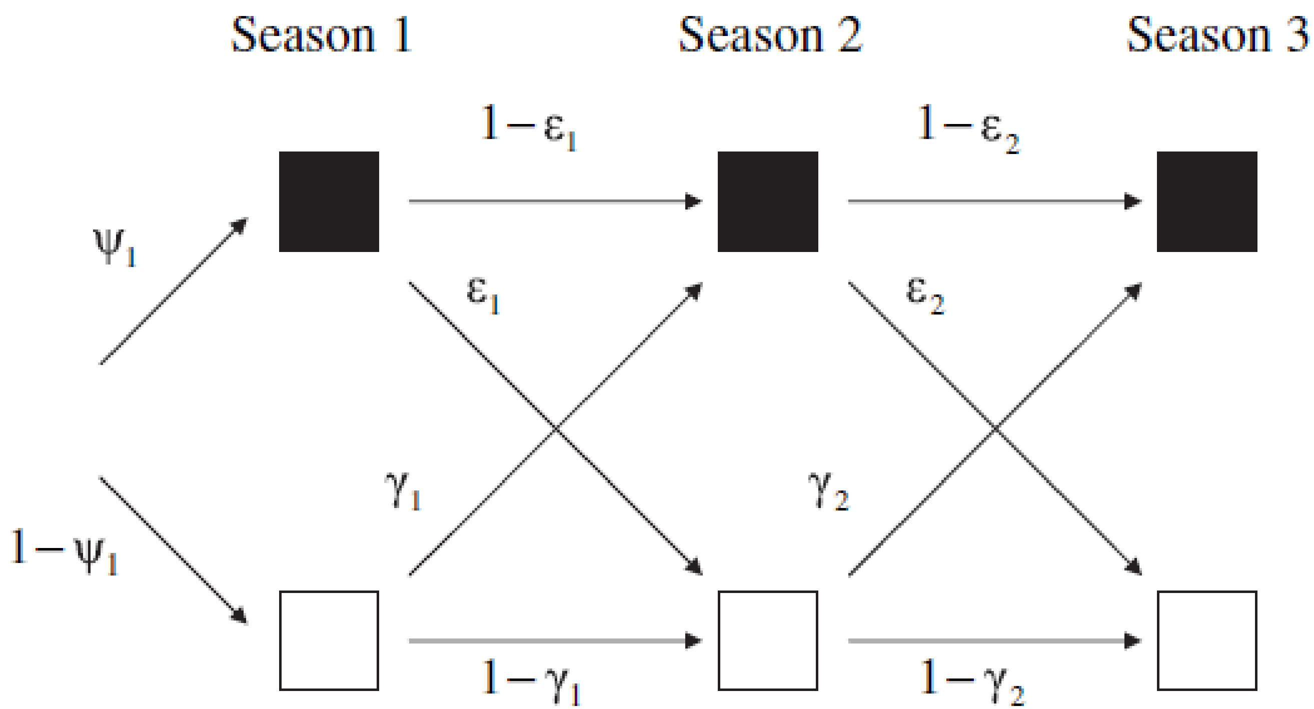


FIGURE 7.2 Representation of how the occupancy state of a site may change between seasons in terms of the processes of occupancy (ψ), colonization (γ), and local extinction (ε). Filled boxes indicate that the site is occupied (species present) in that season, while empty boxes indicate that the site is unoccupied (species absent).

MacKenzie et al. 2006 (book)

Summary – Single season occupancy models

- General, flexible approach that models **occupancy** (ψ) while accounting for imperfect **detection probability** (p)
- Data needs:
    - detection history at sites (need to have repeat sampling at each site!);
    - covariates (if any) – covariates can be for the site OR for the sampling occasion
- Strengths
    - More accurate than assuming p=1
    - Evaluate covariate effects; assess competing hypotheses
    - Flexible, statistical approach
- Drawbacks:
    - Requires more sampling at each site
    - Modeling pres/abs and not abundance
- Extensions
    - Have p depend on abundance; Multiple seasons; multiple species;…
- Know the general branching diagram and be able to derive an equation for 1) the probability for a given detection history, and 2) the likelihood for a simple example