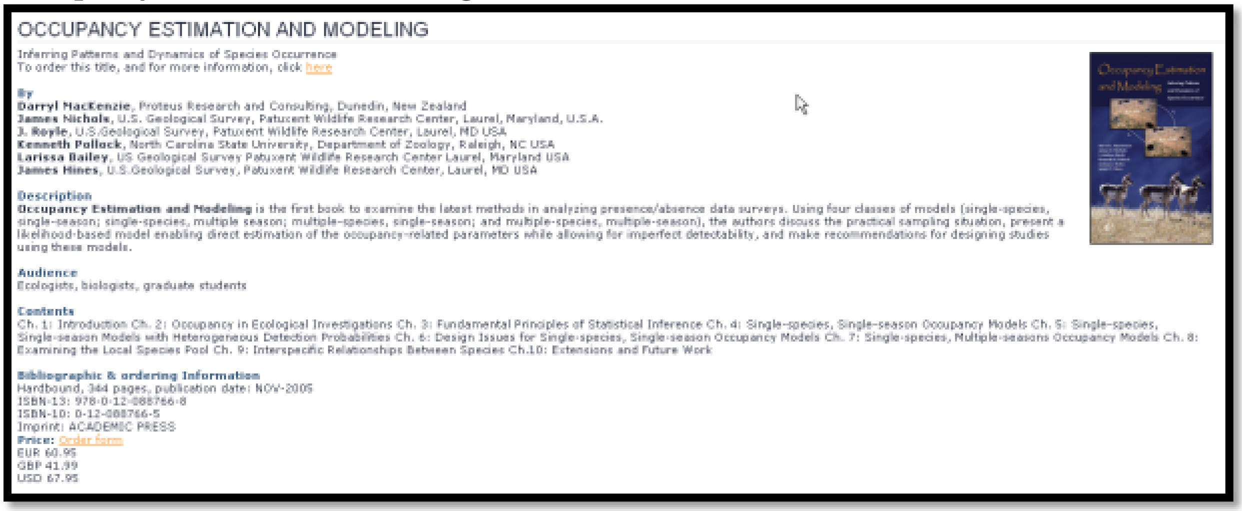
Occupancy estimation & modeling


OCCUPANCY ESTIMATION AND MODELING

Inferring Patterns and Dynamics of Species Occurrence
To order this title, and for more information, click here

By
Darryl MacKenzie, Proteus Research and Consulting, Dunedin, New Zealand
James Nichols, U.S. Geological Survey, Patuxent Wildlife Research Center, Laurel, Maryland, U.S.A.
J. Royle, U.S.Geological Survey, Patuxent Wildlife Research Center, Laurel, MD USA
Kenneth Pollock, North Carolina State University, Department of Zoology, Raleigh, NC USA
Larissa Bailey, US Geological Survey Patuxent Wildlife Research Center Laurel, Maryland USA
James Hines, U.S.Geological Survey, Patuxent Wildlife Research Center, Laurel, MD USA

Description
Occupancy Estimation and Modeling is the first book to examine the latest methods in analyzing presence/absence data surveys. Using four classes of models (single-species, single-season; single-species, multiple season; multiple-species, single-season; and multiple-species, multiple-season), the authors discuss the practical sampling situation, present a likelihood-based model enabling direct estimation of the occupancy-related parameters while allowing for imperfect detectability, and make recommendations for designing studies using these models.

Audience
Ecologists, biologists, graduate students

Contents
Ch. 1: Introduction Ch. 2: Occupancy in Ecological Investigations Ch. 3: Fundamental Principles of Statistical Inference Ch. 4: Single-species, Single-season Occupancy Models Ch. 5: Single-species, Single-season Models with Heterogeneous Detection Probabilities Ch. 6: Design Issues for Single-species, Single-season Occupancy Models Ch. 7: Single-species, Multiple-seasons Occupancy Models Ch. 8: Examining the Local Species Pool Ch. 9: Interspecific Relationships Between Species Ch.10: Extensions and Future Work

Bibliographic & ordering Information
Hardbound, 344 pages, publication date: NOV-2005
ISBN-13: 978-0-12-088766-8
ISBN-10: 0-12-088766-5
Imprint: ACADEMIC PRESS
Price: Order form
EUR 60.95
GBP 41.99
USD 67.95

- Many interesting questions and applied problems in ecology and conservation require data from large spatial scales through a number of years

- Often it is not possible to monitor abundance and/or vital rates across large spatial scales through time

- In such cases it's important to consider state variables other than abundance or population size that may be useful

- Occupancy seems worth considering as a useful state variable. It is defined as the proportion of area, patches, or sample units that is occupied

    - Occupancy may be the state variable of 1st choice in some cases, too, especially where there's interest in changes in rate of occupancy through space and time

        - Studies of distribution and range, e.g., spotted owl, bull trout
        - Relate patch occupancy to patch and/or landscape characteristics
        - Spread of invasion
        - Disease dynamics

    - When working with occupancy, it is crucial to consider the possibility that a species may be present on a site but not detected during a survey. Given this very real possibility, detection still equates to presence, but **non-detection does not necessarily mean absence**

Parameters

- ψ– occupancy probability

- p – detection probability
  - $p^*=1-(1-p)^t$ = prob. detect $\geq$1x in 't' surveys

- Adjust counts of sites where species was observed (s_d) for possible failed detections from repeated surveys of s sites
  - $\hat{\psi} = \frac{s_d}{s \cdot p^*}$

e.g.,

- $p = 0.4$
- $p^* = 1-(1-0.4)^4 = 0.8704$

- $t = 4$
- $s = 100$ sites
- $s_d = 40$ sites

- $\hat{\psi} = \frac{s_d}{s \cdot p^*} = \frac{40}{100 \cdot 0.8704} = 0.46$

- Often questions involve multiple species. Many such questions can be addressed using information from occupancy of sites by multiple species, e.g.,
    - Patterns of species richness in space and time
    - Species assemblages – dependencies among species

A useful analogy can be made between populations and communities (see pg. 556, chapter 20 of *WNC*)

Population of a single species	**Community of multiple species**
Individual animals	Individual species
Abundance of individuals	Species richness
Survival (mortality)	Local persistence (extinction)
Immigration	Colonization

Detection probability is relevant in either case

In either case, we need to take detection into account to properly monitor the quantities of interest.

- Occupancy of area by a single species example
    - $\Pr(detect \mid present, n\ surveys) = 1 - \prod_{i=1}^{n}(1 - p_i)$
    - Might do 3 surveys and miss the species on each occasion
    - E.g., $n=3$, $p_i=0.6$, $p^*=1-(.4)^3=0.936$ … miss 6.4% of time when present
    - E.g., $n=3$, $p_i=0.3$, $p^*=1-(.7)^3=0.657$… miss 34.3% of time when present

- Species richness from counts
    - C_i = number of species counted on time-site i
    - N_i = true number of species present on time-site i
    - $\hat{N}_i = C_i / \hat{p}_i$
        - Let p_{ijk} =probability of detecting individual k of species j on sampling occasion i
        - Species detection probability is: $p_{ij} = 1 - \prod_{k=1}^{n_j}(1 - p_{ijk})$, n_j = abundance of species j
        - Detection is affected by species abundance (less likely to miss all if there are many individuals present), species characteristics (secretiveness, movement patterns, vocalizations), sampling methods used (e.g., observer, gear), sampling conditions (e.g., habitat setting, weather)
    - Evaluating questions based on raw counts without regard to detection probability is not valid unless $p_{ij}=1$ for all species on all occasions. It's hard to imagine that being true in many cases

Models have been developed to deal with 4 broad classes of models:

1. single-species, single-season
2. singles-species, multiple-season
3. multiple-species, single-season
4. multiple-species, multiple-season

There are a number of recent developments that extend these further, e.g., multi-state & multi-scale models.

Allow us to estimate a variety of useful parameters in likelihood framework while allowing for imperfect detectability

1. Single species
    - Single-season: Percent of Area Occupied (PAO)
        1. Detection or non-detection on surveys
        2. Repeat visits used
        3. Estimate PAO with $p < 1$
            - Single state
            - Multi-state
                - E.g., site can be:
                    - unoccupied,
                    - w/ mated pair that produces no young,
                    - w/ mated pair that produces young
            - Multi-scale

    - Multiple-season: Time-specific rates of:
        1. Occupancy
        2. Local rates of extinction and colonization – think about how would these be affected if ignore *p* and $p<1$?
            - E.g., probability of extinction or colonization as functions of patch size, isolation, connectivity
            - Equilibrium occupancy rates
        3. Variation in rates as function of covariates (patch features, year features, management actions)

    - Applications
        1. Range
        2. Habitat relationships – *p* may vary with habitat setting
        3. Metapopulation dynamics
        4. Broad-scale monitoring – usually cheaper to estimate occupancy than abundance
        5. Assess conservation status and changes in status

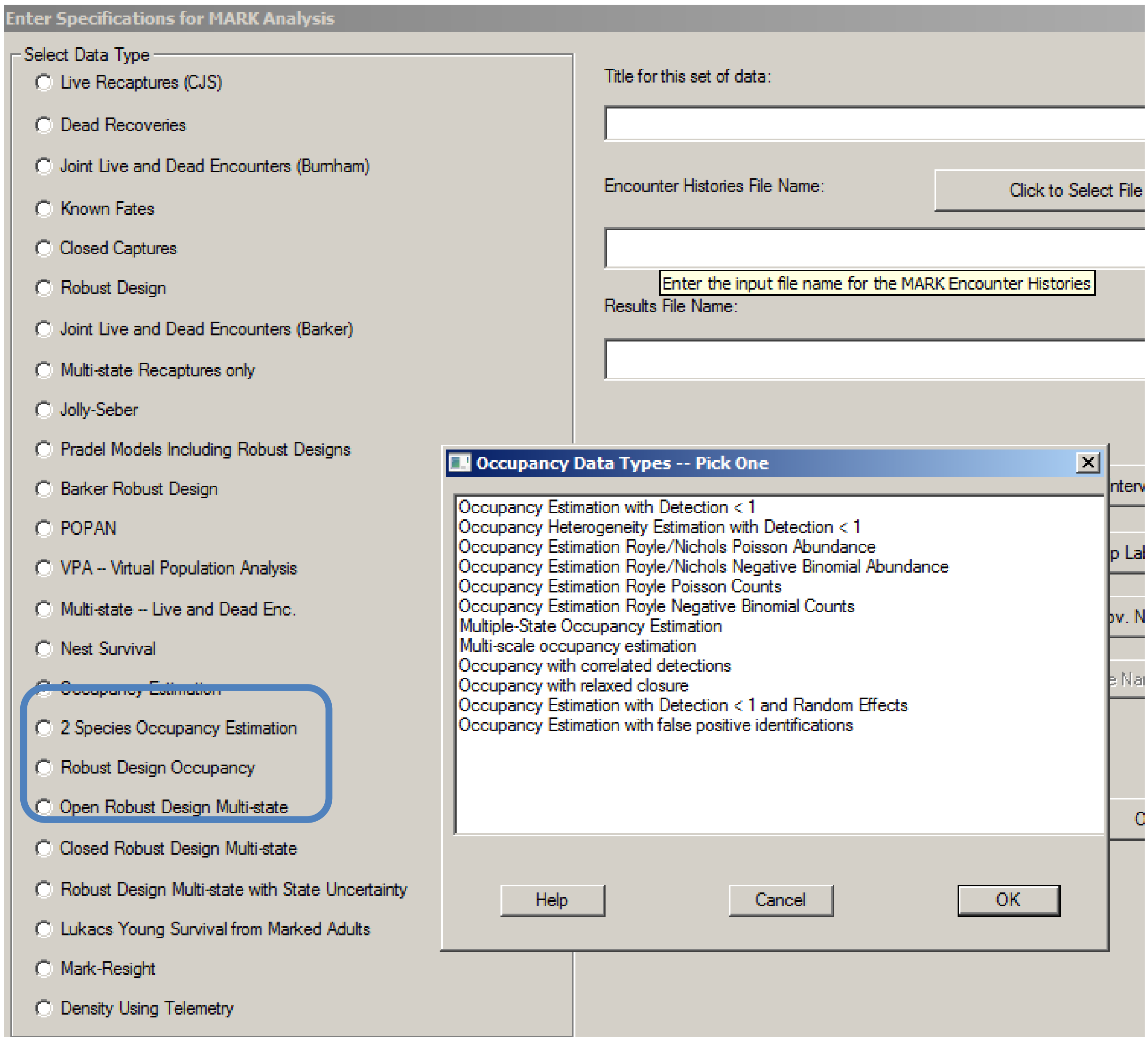
2. Multiple species
    - # of species present on sites – species richness
    - Community similarity among sites
    - Dynamics - species turnover rates
    - Colonization and extinction rates as functions of P/A for other species
    - Species interactions
    - Much work done on 2 species situation
        - Example: northern spotted owl & barred owl
            - Does occupancy of site by 1 species depend on P/A of the other
            - Does detection of site by 1 species depend on P/A of the other
            - Does detection of site by 1 species depend on detection of the other

Lots of Applications in recent literature & expect MANY more to come soon

- Amphibian occupancy of sites relative to site characteristics
- Bird occupancy of sites relative to site characteristics,
    - e.g., spotted owls, songbirds, colonial nesters
- Amphibian community dynamics
- BBS data & forest fragmentation effects on communities

Running these types of models

- MARK has occupancy models for single species and 2-species models. There are 2 versions explicitly designed to handle heterogeneity in *p* not associated with measured covariates (see the 2nd and 2nd to last of the Occupancy Data Types in the figure below)



- Specialized software exists to handle a variety of situations, e.g., Program PRESENCE, R packages ‘unmarked’ and ‘RPresence’.

- Basic approach is not too big a leap for you as likelihoods look familiar
    - ψ = probability of site being occupied
    - p = probability of being detected given presence

Getting Started: one species - one season

- 2 processes involved: occupancy & detection
  - If not occupied, then no possibility for detection
  - If occupied, then can detect or not on each occasion of sampling
- $\Pr(eh_i=0101) = \psi(1-p_1)p_2(1-p_3)p_4$
- $\Pr(eh_i=0000) = \psi\prod_{j=1}^{4}(1-p_j)+(1-\psi)$

$$L(\psi, p \mid eh_1, eh_2, ..., eh_3) = \prod_{i=1}^{s}\Pr(eh_i) = \left[\psi^{s_D}\prod_{j=1}^{K}p_j^{s_j}(1-p_j)^{s_D-s_j}\right]\left[\psi\prod_{j=1}^{K}(1-p_j)+(1-\psi)\right]^{s-s_D}$$

where s is # of sites, s_D is # of sites where the species was detected at least once, and s_j is # of sites where the species was detected during the jth survey.

Assumptions

1. Occupancy state of sites is constant during all single-season surveys
2. Probability of occupancy is equal across all sites
3. Probability of detection given occupancy is equal across all sites
4. Detection of species in each survey of a site is independent of those on other surveys
5. Detection histories at each location are independent

Can relax these assumptions and use covariate modeling. Such modeling will often be of primary interest in an occupancy study.

Example:

- Data for Blue-Ridge two-lined salamander from surveys in 2001 in Great Smoky Mountains National Park (see page 99 of MacKenzie et al. 2006 book)

Transect	Occ 1	Occ 2	Occ 3	Occ 4	Occ 5
1	0	0	0	1	1
2	0	1	0	0	0
3	0	1	0	0	0
...					
39	0	0	0	0	1

- Naïve estimates = $n_{detected}/n_{sampled}$ = 18/39 = 0.46

- Estimates from *Psi(.),p(.)* model

```
                     Real Function Parameters of {p(t),psi(.)}
                                                          95% Confidence Interval
 Parameter                  Estimate       Standard Error      Lower           Upper
-------------------------  --------------  --------------  --------------  ----------
    1:p                     0.2587291       0.0577019       0.1621557       0.3862995
    2:Psi                   0.5946226       0.1225985       0.3512006       0.7989882
```

- Use estimates of *p(.)* to estimate probability of missing salamanders on occupied site on all 5 surveys: $(1\text{-}0.26)^5 = 0.224$. So, expect to miss them on 22.4% of occupied sites and to detect them on 77.6% of sites. And, 0.4615/.5946 = 0.776

- Can use the estimates of *psi* & *p(.)* to estimate the probability that a site is occupied given that no salamanders were detected there in 5 surveys. This is worked out on page 100 of MacKenzie et al. (2006)

- Estimates from *Psi(.),p(t)* model

Real Function Parameters of {p(t),psi(.)}

Parameter	Estimate	Standard Error	95% Confidence Interval Lower	Upper
1:p	0.1768717	0.0845125	0.0644377	0.4013302
2:p	0.1326537	0.0740540	0.0415185	0.3506508
3:p	0.3979612	0.1190041	0.1998064	0.6363531
4:p	0.3537433	0.1137000	0.1711581	0.5919885
5:p	0.2653075	0.1010181	0.1156440	0.4993047
6:**Psi**	**0.5798786**	0.1175670	0.3489652	0.7804244

Covariates:
Can evaluate whether *psi*¸ *p*, or both are best modeled as functions of covariates. E.g., in MacKenzie et al. (2006):

$\hat{\psi} = \exp(0.02 + 1.17 browse) / (1 + \exp(0.02 + 1.17 browse))$, where *browse* is an indicator variable telling whether or not a site was browsed or not by animals

$$\hat{\psi}_{unbrowsed} = \exp(0.02) / (1 + \exp(0.02)) = 1.02 / 2.02 = 0.50$$
$$\hat{\psi}_{browsed} = \exp(1.19) / (1 + \exp(1.19)) = 3.29 / 4.29 = 0.77$$

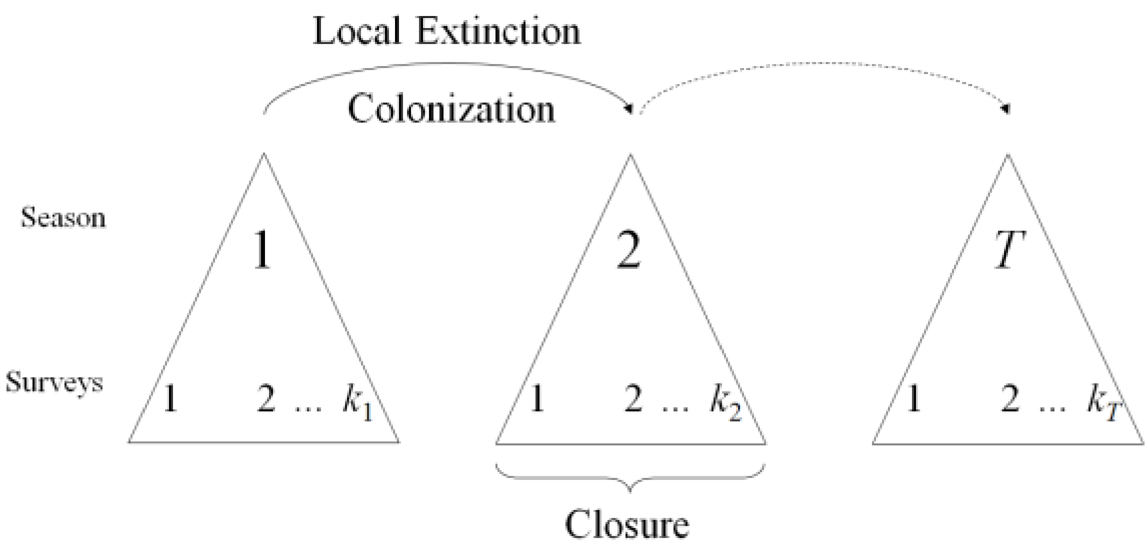
And, it's often useful to model *p* as a function of covariates as well. For example, detection rate might differ among observers, by weather, by habitat conditions, and other factors. *There is an excellent example with covariates in Chapter 21 of C&W by Gerber, Mosher, Martin, Bailey, and Chambert* (see section 21.1.3).

Heterogeneity in *p*

It is quite possible that *p* varies among sites for a variety of reasons and that you may not have the covariates that reflect that variation, e.g., *p* varies due to variation in abundance of the species at different sites. We have many of the same problems and solutions available to us that we encountered when working with closed models.

Single species – multiple seasons

- Occupancy dynamics through time
- Model changes in occupancy over time (e.g., from one year to the next)
- Use multiple surveys per year across multiple years where occupancy does not change w/in a year but might change across years. Remember robust design for mark-recapture data and imagine an analogue for occupancy studies.
- Parameters
  - ψ_t = occupancy rate by year
  - $\lambda_t = \psi_{t+1} / \psi_t$ = rate of change in occupancy
  - ε_t = P(absence at time *t*+1 | presence at *t*) = patch extinction probability
  - γ_t = P(presence at *t*+1 | absence at *t*) = patch colonization probability



- History : <u>10</u> 00 11 0<u>1</u>
  - ↑*primary$_i$* ↑*secondary$_j$*
- 10, 11, 01 = known presence (assume no false detections)
- Interior ‘00’ =
  - Site occupied but occupancy not detected, *or*
  - Site unoccupied (in this case = locally extinct & recolonized later)

Probability for this history w/ time-varying parameters and standard parameterization, which works with $\psi_1, \varepsilon_t, \gamma_t$

$$\Pr(10\ 00\ 11\ 01) = \psi_1 p_{11}(1-p_{12})[(1-\varepsilon_1)(1-p_2^*)(1-\varepsilon_2)+\varepsilon_1\gamma_2]p_{31}p_{32}(1-\varepsilon_3)(1-p_{41})p_{42}$$

$$\Pr(\text{occupied at } t_2 \mid \psi_1, \varepsilon_1, \gamma_1) = \psi_2 = \psi_1(1-\varepsilon_1)+(1-\psi_1)\gamma_1$$

Can also parameterize in terms of: (ψ_1, ε_t, λ_t), (ψ_1, λ_t, γ_t), (ψ_t, ε_t), or (ψ_t, γ_t)

Within each season, model ψ, ε, γ, p, which must be constant w/in a season but can vary by season and be a function of covariates (e.g., patch size, patch isolation, habitat features)

For each survey, model p, which can vary among surveys with features such as observers, environmental conditions, etc.

Main Assumptions:

1) Patches are independent with respect to site dynamics and identifiable
   - Independence violated when sub-patches exist within a site
2) No colonization and extinction among secondary periods
   - Can be violated if patches are settled or disappear across secondary periods due to features such as arrival/departure for migratory species, disturbances
3) Patches have identical colonization and extinction rates or heterogeneity in rates is adequately modeled with identified patch covariates
   - Violated with unidentified heterogeneity & reduced via stratification

<u>Working with output from Multi-season Occupancy Models</u>

Estimating $\hat{\psi}_{t+1}$ from estimates of $\hat{\psi}_t$, $\hat{\gamma}_t$, and $\hat{\varepsilon}_t$

Example: $\hat{\psi}_1 = 0.8,\ \hat{\gamma}. = 0.4,\ \hat{\varepsilon}. = 0.2$

$$\hat{\psi}_{t+1} = \hat{\psi}_t \cdot (1-\hat{\varepsilon}_t) + (1-\hat{\psi}_t)\cdot\hat{\gamma}_t$$
$$\hat{\psi}_2 = 0.8\cdot(1-0.2) + (1-0.8)\cdot 0.4$$
$$\hat{\psi}_2 = 0.64 + 0.08 = 0.72$$

$$\hat{\psi}_3 = 0.72\cdot(1-0.2) + (1-0.72)\cdot 0.4$$
$$\hat{\psi}_3 = 0.576 + 0.112 = 0.688$$

$$\hat{\psi}_4 = 0.688\cdot(1-0.2) + (1-0.688)\cdot 0.4$$
$$\hat{\psi}_4 = 0.5504 + 0.1248 = 0.6752$$

If the colonization and extinction values are constant, the system's occupancy rate will eventually reach an equilibrium. The equilibrium level is the occupancy rate for which the number of colonization events equals the number of extinction events. In terms of equations, $\psi_{equilib}$ occurs when $(1-\psi)\cdot\gamma = \psi\cdot\varepsilon$. To find the equilibrium value, we can work with the equations and do some re-arrangement of terms.

$$(1-\psi)\cdot\gamma = \psi\cdot\varepsilon$$
$$\gamma - \gamma\cdot\psi = \varepsilon\cdot\psi$$
$$\gamma = \gamma\cdot\psi + \varepsilon\cdot\psi$$
$$\gamma = (\gamma+\varepsilon)\cdot\psi$$
$$\frac{\gamma}{\gamma+\varepsilon} = \psi$$

$$\psi_{equilib} = \frac{0.4}{0.4+0.2} = 0.6667$$

When $\psi = 0.6667$, extinctions offset colonizations, and each year
~13.3% of sites will go extinct &
~13.3% of sites will be colonized.

Resources for learning, designing, conducting, and analyzing occupancy studies

Check out: https://www.mbr-pwrc.usgs.gov/software/presence.html for lots of excellent resources including important literature and software. At that site, you can find downloads of PRESENCE, RPresence, and Genpres, which is a program that you can use to generate data and analyze it in Programs MARK or PRESENCE. Such software can be very helpful in designing studies.