Design and Analysis of Occupancy Studies Part 1c

Carl James Schwarz

Department of Statistics and Actuarial Science
Simon Fraser University
Burnaby, BC, Canada
cschwarz @ stat.sfu.ca

Outline I

Single Species; Single-Season - Planning

Planning and Study Design

Single Species; Single-Season - Planning

Starting principles:

- Randomization - makes your sample representative of population
- Replication - controls precision
- Stratification - controls for noise, e.g. via covariates

No amount of statistical wizardry can rescue a badly executed survey.

Single Species; Single-Season - Planning

Starting principles:

- Must have some sites visited more than once to estimate detection probability
- Objectives (can) change study design:
    - Objective: Estimate overall occupancy proportion. Design: Select sites completely at random from relevant population. No control on number of sites in habitat types, so comparison of occupancy rates between habitat types may have poor power because of small number in one habitat type
    - Objective: Compare occupancy rates between habitats. Design: Stratify by habitat type, select equal number of sites from each habitat type. If you want overall occupancy, you must weight habitat occupancy by proportion of entire population of sites in each habitat type.
- Occupancy $\neq$ Resource Selection.

Single Species; Single-Season - Planning

Defining a "site":

- What is spatial scale were presence/absence is meaningful?
  - E.g. Remnant forest stands and rare species. Is information at stand level sufficient or do you need information on what fraction of stand is occupied?
  - E.g. Large vs. small home ranges.
- "Larger" sites have higher probabilities of occupancy than "smaller" sites (*ceteris paribus*).
  - Rule of thumb: occupancy should be $0.2 \rightarrow 0.8$ over your sites.

Single Species; Single-Season - Planning I

Site selection

- Randomize, randomize, randomize !
  - ONLY time non-random sampling acceptable is a census.
- Methods of this class assume Simple Random Sample
  - Each site has EQUAL probability of selection
  - Each site selected independently of other sites
- More complex designs possible, but beyond scope of this course (and current software)
  - Sites are forest stands of different areas and selection is proportion to size of stand (pps).
  - Sites are selected adaptively in waves, i.e. sites near where occupancy found are selected with higher probability.
  - Sites are selected using cluster and multi-state designs, e.g. random select stands, measure all trees in stand.

Single Species; Single-Season - Planning II

- AVOID selecting sites based on "prior" knowledge of occupancy UNLESS you are only interested in changes in occupancy of these sites. E.g. Selecting previously occupied stands and measuring change in occupancy over time in multi-season models.

Single Species; Single-Season - Planning

Defining a "Season".

- Critical assumption of closure, i.e. occupancy of site does not change over season.
  - Random movement is analyzable but interpretation of "occupancy" must be modified.
  - Immigration/emigration lead to estimates of occupancy with no direct interpretation.
- Length of season depends on stability of population, i.e. slowly moving animals can be survey over longer "seasons" with closure among sites.
- "Larger" sites can have longer "Seasons" as closure more likely to be satisfied for slowly moving animals, but local deaths may be problematic.

Single Species; Single-Season - Planning I

Conducting repeat surveys.

- Many options available:
  - Visit site multiple times with a single survey per visit.
  - Visit site one with multiple INDEPENDENT surveys (e.g. different observers, different transects, different quadrats to look for fecal pellets)
- Key is that repeat surveys need to be INDEPENDENT
  - CAUTION: Detect an animals den on first visit; second visit keys on den location.
    - Use different observer on each visit who does not know location of den.
    - Use "removal" method (see later) where surveys stop after occupancy established.
    - Define "already detected" covariate in modelling

Single Species; Single-Season - Planning II

- CAUTION: Multiple simultaneous surveys with very low density (e.g. one nest per site and several transects are run) are PROBLEMATIC because if one survey detects the nest, the other survey MUST (by definition) not detect the nest.
- How will you align different surveys if models with $p(t)$ are used (e.g. think of the American Toad exercise).

Single Species; Single-Season - Planning

Conducting repeat surveys (continued).

- Avoid confounding observer/ site/ temporal effects.

Design A

Site	Day 1			Day 2			Day 3		
1	X	X	X						
2							X	X	X
3				X	X	X			
4				X	X	X			
5							X	X	X
6	X	X	X						
7				X	X	X			
8							X	X	X
9	X	X	X						
p	0.5			0.3			0.8		

Design B

Site	Day 1			Day 2			Day 3		
1	X				X				X
2			X	X			X		
3		X				X		X	
4		X		X					X
5			X			X	X		
6	X				X				X
7		X		X				X	
8			X		X		X		
9	X					X		X	
p	0.5			0.3			0.8		

Design B is better. [From MacKenzie et al (2006)]

- Rotate observers among sites and surveys to avoid consistent observer effects.

Single Species; Single-Season - Planning

Allocation of effort: Number of sites vs. Number of surveys.
MacKenzie and Royle (2005). J. Applied Ecology, 42, 1105-1114.
doi: 10.1111/j.1365-2664.2005.01098.x

Need to know:

- Level of acceptable precision for occupancy estimate
  - Preliminary survey: SE of 25% of $\widehat{\psi}$, i.e. if $\widehat{\psi} = 0.80$ then $se \approx 0.20$.
  - Management work: SE of 10% of $\widehat{\psi}$, i.e. if $\widehat{\psi} = 0.80$ then $se \approx 0.08$.
  - Scientific work: SE of 5% of $\widehat{\psi}$, i.e. if $\widehat{\psi} = 0.80$, then $se \approx 0.04$.
- Initial guess of probability of occupancy and detection. (Past surveys; other similar work).
- Resources available (e.g. what is your budget).

Single Species; Single-Season - Planning

MINIMUM number of sites needed:

Assume detectability = 100% in each site.

$$SE(\widehat{\psi})_{100\% detectability} = \sqrt{\frac{\psi(1-\psi)}{s}}$$

where s is the number of sites. E.g. if $\psi \approx 0.8$ and have $s = 100$, then

$$SE(\widehat{\psi})_{100\% detectability} = \sqrt{\frac{0.8(0.2)}{100}} = 0.04$$

which is acceptable for scientific work.
BUT DETECTABILITY < 100% so this is a LOWER bound on actual SE.

Single Species; Single-Season - Planning

Standard design with s surveyed sites and K surveys per site.
Total number of surveys $TS = s \times K$.

$$SE(\widehat{\psi}) = \sqrt{\frac{\psi}{s}\left[(1-\psi) + \frac{1-p*}{p*-Kp(1-p)^{K-1}}\right]}$$

where $p* = 1-(1-p)^K$ is the probability of detecting species at least once in K surveys if the site is occupied.

See the spreadsheet that comes with notes.

Single Species; Single-Season - Planning

For example, what precision for $\widehat{\psi}$ would be obtained when $\psi \approx 0.8$, $p \approx 0.4$, $s = 85$ and $k = 6$?

0.8	Approximate occupancy rate (psi)
0.4	Approximate detection on each SURVEY (p)
50	Number of sites (s)
6	Number of surveys/site (K)
0.95	p* (probability of detection on a site over all K surveys)
0.0646	Estimated SE of psi-hat

Notice that $p* = .95$ so that very few sites will be false negatives.

Single Species; Single-Season - Planning

Optimal number of surveys/site (ignoring costs) for standard design:

	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

Source: MacKenzie and Royle (2005).

For very low detectability, need to take many surveys in each site!

Single Species; Single-Season - Planning

Suppose you have enough budget for 500 surveys and $\psi = 0.8$ and $p = 0.4$?

Use $K = 6$ (from table) which implies $s = 500/6 = 85$ and spreadsheet gives:

0.8	Approximate occupancy rate (psi)
0.4	Approximate detection on each SURVEY (p)
85	Number of sites (s)
6	Number of surveys/site (K)
0.95	p* (probability of detection on a site over all K surveys)
0.0495	Estimated SE of psi-hat

Is this precise enough?

Single Species; Single-Season - Planning

Typically costs of finding a new site (c_{new}) are $>>$ cost of each survey in a site (c_{survey}). Total cost of s sites and K surveys per site is then

$$C = s \times c_{new} + sK \times c_{survey}$$

You can use the same workbook to find the optimal choice using solver subject to constraints.

Single Species; Single-Season - Planning

Suppose $\psi \approx 0.8$, $p \approx 0.4$, $c_{new} = 10$ (hours) and $c_{survey} = 1$ (hour). The budget allowable is 2000 (hours). What is the optimal tradeoff?

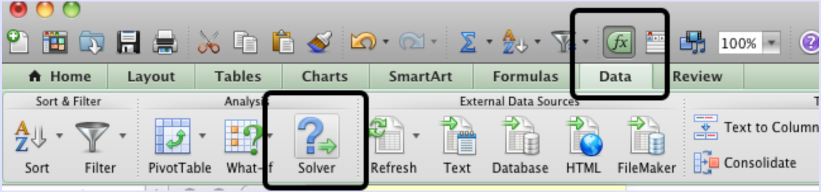
Enter basics in start of spreadsheet with a preliminary allocation of sites and surveys/site based on previous tables and allowable budget:

12	0.8	Approximate occupancy rate (psi)
13	0.4	Approximate detection on each SURVEY (p)
14		
15	10	Cost for a new site (c_new). Just RELATIVE COSTS are important
16	1	Cost per survey (c_survey)
17	2000	Maximum allowable cost
18		
19	100	Number of sites (s)
20	6	Number of surveys/site (K)
21		
22	0.95	p* (probability of detection on a site over all K surveys)
23	1600.00	Total cost
24		
25	0.0457	Estimated SE of psi-hat

Single Species; Single-Season - Planning

Suppose $\psi \approx 0.8$, $p \approx 0.4$, $c_{new} = 10$ (hours) and $c_{survey} = 1$ (hour). The budget allowable is 2000 (hours). What is the optimal tradeoff?

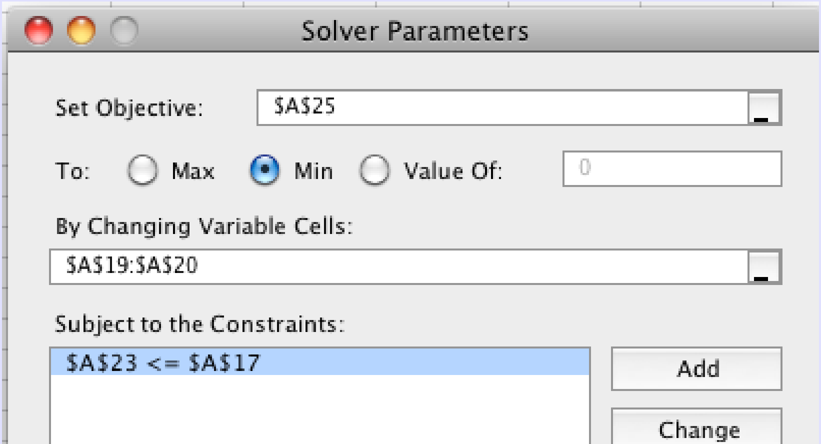
Select the solver:



Single Species; Single-Season - Planning

Suppose $\psi \approx 0.8$, $p \approx 0.4$, $c_{new} = 10$ (hours) and $c_{survey} = 1$ (hour). The budget allowable is 2000 (hours). What is the optimal tradeoff?

Complete solver dialogue box:



Single Species; Single-Season - Planning

Suppose $\psi \approx 0.8$, $p \approx 0.4$, $c_{new} = 10$ (hours) and $c_{survey} = 1$ (hour). The budget allowable is 2000 (hours). What is the optimal tradeoff?

Save solutions

0.8	Approximate occupancy rate (psi)
0.4	Approximate detection on each SURVEY (p)
10	Cost for a new site (c_new). Just RELATIVE COSTS are important
1	Cost per survey (c_survey)
2000	Maximum allowable cost
112.963208	Number of sites (s)
7.70487963	Number of surveys/site (K)
0.98	p* (probability of detection on a site over all K surveys)
2000.00	Total cost
0.0397	Estimated SE of psi-hat

Single Species; Single-Season - Planning

Some general principles:

- Low detectability on each survey $\rightarrow$ LOTS of surveys/site!
- Low occupancy or high occupancy require fewer sites than intermediate levels of occupancy (easy to estimate 0 or 100%).
- Low occupancy implies fewer surveys/site; higher occupancy implies more surveys/site.
- Cheap surveys imply more surveys and fewer sites.
- NUMBER of SITES is the PRIMARY factor in success.

Aim to get $p(detection|present)$ to be 0.80 or higher (see next few slides for details).

Single Species; Single-Season - Planning

Alternative designs: Survey some sites intensively, some sites **only once**.

Idea is that sites with many surveys estimate p and after a point, additional information is not useful and better to survey more sites. NOT RECOMMENDED - See MacKenzie et al (2006).

- Unless $p > .8$ no benefit.
- Typically $c_{new} >> c_{survey}$ which negates taking more sites measured only once.
- May be of use when some sites are remote and dangerous to access.

Single Species; Single-Season - Planning

Alternative designs: "Removal" method.

Idea is to stop surveying sites after occupancy is (positively) confirmed. Can be more efficient that standard design

- Once species confirmed present, can you shift resources to more sites?
- Are you prepared to survey some sites longer than under standard design (uncertainty in planning)?
- Do repeat surveys use knowledge of occupancy (e.g. den) to revisit in next survey (not independent).
- CAUTION: Must assume that $p_i = p_j$ for at least one pair of (i, j) to fit a $p(t)$ model.
- Hybrid designs where some sites are surveyed completely, and some sites are surveyed until occupancy confirmed.

Single Species; Single-Season - Planning

Optimal number of (max) surveys/site for removal method.

TABLE 6.5 Optimal Maximum Number of Surveys to Conduct at Each Site for a Removal Design (K) Where All Sites Are Surveyed Until the Species Is First Detected, for Selected Values of Occupancy (ψ) and Detection Probabilities (p)

	ψ								
p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	23	24	25	26	28	31	34	39	49
0.2	11	11	12	13	13	15	16	19	23
0.3	7	7	7	8	8	9	10	12	14
0.4	5	5	5	6	6	6	7	8	10
0.5	4	4	4	4	4	5	5	6	8
0.6	3	3	3	3	3	4	4	5	6
0.7	2	2	2	3	3	3	3	4	5
0.8	2	2	2	2	2	2	3	3	4
0.9	2	2	2	2	2	2	2	2	3

Source: MacKenzie and Royle (2005).

Compare to previous table (slide 14).

Single Species; Single-Season - Planning

Efficiency of optimal "removal" design to standard design with equal total effort.

TABLE 6.6 Ratio of Standard Errors for Optimal Standard and Removal Designs

	ψ								
p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.90	0.94	0.98	**1.04**	**1.10**	**1.18**	**1.30**	**1.46**	**1.74**
0.2	0.91	0.94	0.99	**1.04**	**1.10**	**1.18**	**1.28**	**1.44**	**1.71**
0.3	0.92	0.95	0.99	**1.04**	**1.10**	**1.17**	**1.27**	**1.42**	**1.68**
0.4	0.93	0.96	0.99	**1.03**	**1.09**	**1.17**	**1.26**	**1.40**	**1.64**
0.5	0.93	0.96	1.00	**1.04**	**1.08**	**1.16**	**1.24**	**1.37**	**1.60**
0.6	0.94	0.97	**1.01**	**1.06**	**1.09**	**1.15**	**1.22**	**1.35**	**1.55**
0.7	0.95	0.96	0.97	**1.01**	**1.07**	**1.13**	**1.22**	**1.31**	**1.48**
0.8	1.00	**1.02**	**1.04**	**1.07**	**1.09**	**1.11**	**1.15**	**1.25**	**1.45**
0.9	**1.02**	**1.05**	**1.07**	**1.10**	**1.13**	**1.17**	**1.20**	**1.24**	**1.31**

Values greater than 1 (in bold) indicate situations where an optimal removal design has a smaller standard error than the optimal standard design.
Source: MacKenzie and Royle (2005).

No real gain/loss unless occupancy high.

Single Species; Single-Season - Planning

General remarks:

- Increasing number sites at cost of number of surveys may not be optimal. E.g. compare SE($\widehat{\psi}$) when $\psi = 0.4$, $p = 0.3$ with $(s = 200, K = 2)$ vs. $(s = 80, K = 5)$.
- Try and reduce probability of false negative when the site is occupied to $0.05 \rightarrow 0.15$, i.e. choose K such that $0.05 < (1-p)^K < 0.15$.
- Optimal $\neq$ robust, i.e. violations of assumptions (e.g. heterogeneity) can cause major problems. The Standard Design is more robust than Removal Design to violations of assumptions. If heterogeneity is problem, recommend that $K \geq 3$.
- Rare species $\rightarrow$ More sites and fewer surveys.
- To investigate hybrid designs/ robustness, simulation approach needed.
- RUN A PILOT STUDY!

Single Species; Single-Season - GENPRES

GENPRES: A planning tool for occupancy studies

This tool allows you simulate effects of violations of assumptions, hybrid designs, or study size on the precision and bias of the estimates.

Useful for single season/multiple season; single species/ two species; etc.
Can do a bootstrap goodness-of-fit test from an analysis.
But . . . sometimes you need to combine GENPRES and PRESENCE/MARK (see multi-season models).

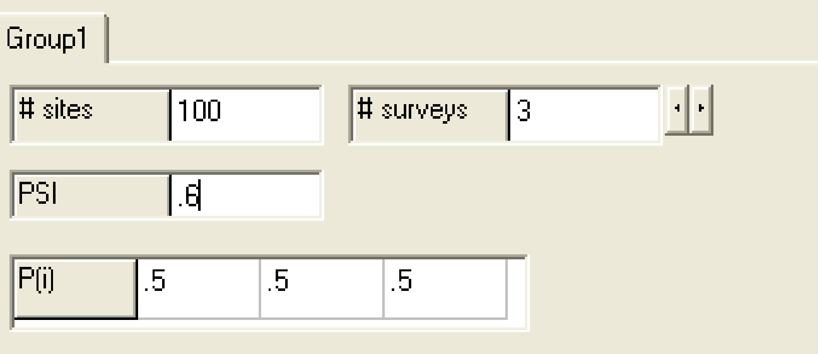
Single Species; Single-Season - GENPRES

Launch GENPRES; make sure Single Species; Single Site model is selected:

Single Species; Single-Season - GENPRES

What are the precision and bias if I do $s = 100$ sites and $K = 3$ with $\psi \approx 0.6$, and $p \approx 0.5$?

Enter these values in the tabbed window:



Single Species; Single-Season - GENPRES

What are the precision and bias if I do $s = 100$ sites and $K = 3$ with $\psi \approx 0.6$, and $p \approx 0.5$?

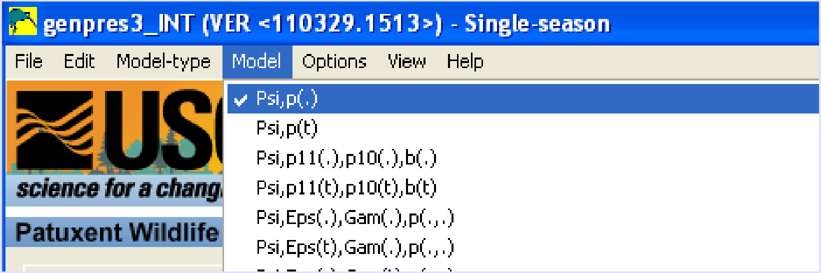
Two methods for studying design:

- Generate EXPECTED counts for each detection history. Analyze EXPECTED counts as data. [Faster and usually sufficient, see Devineau et al (2005)], but in some cases too difficult to do (heterogeneity among animals etc).
- Generate multiple sets of simulated data, analyze each set of simulated data, look at mean/std of estimates.

Single Species; Single-Season - GENPRES

What are the precision and bias if I do $s = 100$ sites and $K = 3$ with $\psi \approx 0.6$, and $p \approx 0.5$?

Select the model with which the data will be analyzed:



Press the *Expected Value* button

Single Species; Single-Season - GENPRES

What are the precision and bias if I do $s = 100$ sites and $K = 3$ with $\psi \approx 0.6$, and $p \approx 0.5$?

```
============================================================

   Individual Site estimates of <Psi>
        Site          Survey          Psi      Std.err     95% conf. interval
   1        site_1          survey_1:  0.6000   0.0624      0.4739 - 0.7141

============================================================

   Individual Site estimates of <p>
        Site          Survey           p       Std.err     95% conf. interval
   1        site_1          survey_1:  0.5000   0.0493      0.4046 - 0.5954
   1        site_1          survey_2:  0.5000   0.0493      0.4046 - 0.5954
   1        site_1          survey_3:  0.5000   0.0493      0.4046 - 0.5954

============================================================
```

Estimates look to be unbiased with expected SE given. Compare to Excel Workbook results.

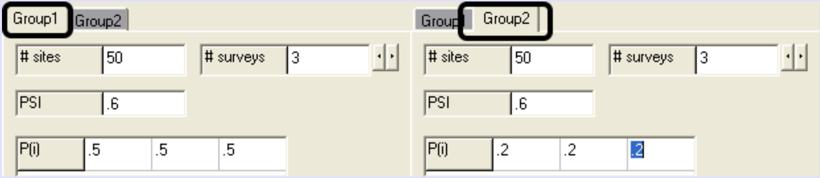
Single Species; Single-Season - GENPRES

What are the precision and bias if I do $s = 100$ sites and $K = 3$ but hidden heterogeneity in detectability (e.g. due to size, breeding status, coloration, etc). We now create two (or more) groups with different parameters in the approximate ratio in the population.

Group 1: $\psi \approx 0.6$, and $p \approx 0.5$ in 50 sites

Group 2: $\psi \approx 0.6$, and $p \approx 0.2$ in 50 sites

Add a group and enter the individual group parameters.



Single Species; Single-Season - GENPRES

Group 1: $\psi \approx 0.6$, and $p \approx 0.5$ in 50 sites
Group 2: $\psi \approx 0.6$, and $p \approx 0.2$ in 50 sites

```
======================================================================

   Individual Site estimates of <Psi>
        Site          Survey            Psi      Std.err      95% conf. interval
   1        site_1          survey_1:  0.5174    0.0742       0.3745 - 0.6575

======================================================================

   Individual Site estimates of <p>
        Site          Survey              p      Std.err      95% conf. interval
   1        site_1          survey_1:  0.4059    0.0584       0.2983 - 0.5233
   1        site_1          survey_2:  0.4059    0.0584       0.2983 - 0.5233
   1        site_1          survey_3:  0.4059    0.0584       0.2983 - 0.5233

======================================================================
```

Are the estimates meaningful?

Single Species; Single-Season - Planning - EXERCISE I

Total budget is \$500,000. Cost per new site is \$2,000. Cost per survey is \$400.
$\psi \approx .4$, $p \approx .3$.

Plan a standard design. What is forecasted precision on $\widehat{\psi}$?

Single Species; Single-Season - Planning - EXERCISE II

What is the impact of going to a hybrid design, i.e. compare

Standard design, 24 sites, 192 surveys

	Survey							
# sites	1	2	3	4	5	6	7	8
24	x	x	x	x	x	x	x	x

Panel design, 36 sites, 144 surveys

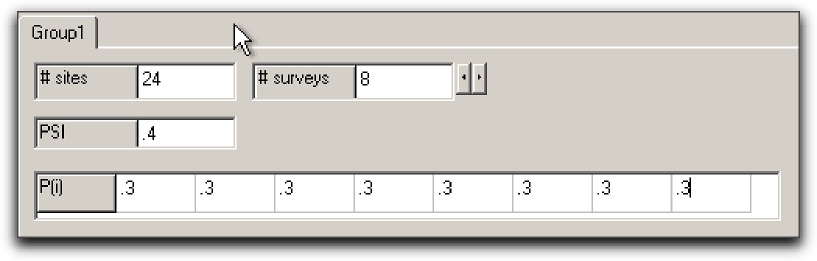
	Survey							
# sites	1	2	3	4	5	6	7	8
12	x	x	x	x	x	x	x	x
6	x	x	-	-	-	-	-	-
6	-	-	x	x	-	-	-	-
6	-	-	-	-	x	x	-	-
6	-	-	-	-	-	-	x	x

$\psi \approx .4$, $p \approx .3$. Compare results of $p(*)$ and $p(t)$ models.

Single Species; Single-Season - Planning - EXERCISE II

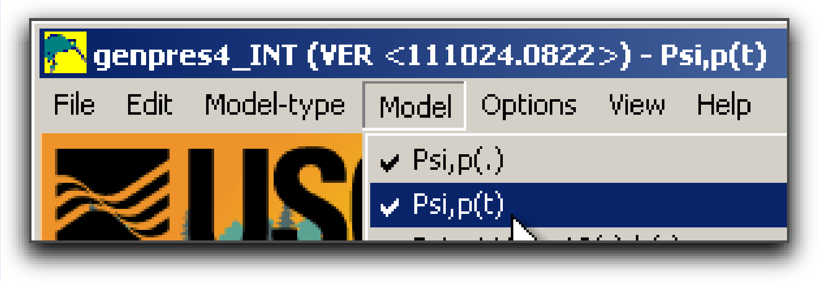
Standard design, 24 sites, 192 surveys								
	Survey							
# sites	1	2	3	4	5	6	7	8
24	x	x	x	x	x	x	x	x

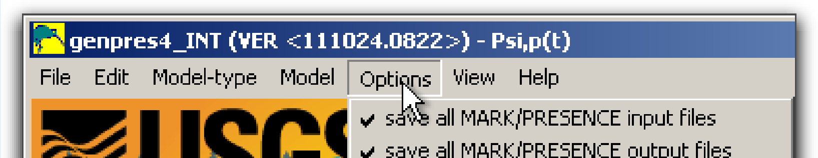
Launch GENPRES, select generating model, and complete:



Single Species; Single-Season - Planning - EXERCISE II

Select which model to analyze data with (you can specify more than one but it becomes messy quickly), and save all of the output (able to reproduce everything). Don't forget to save your simulation settings using File → Save.





Single Species; Single-Season - Planning - EXERCISE II

Scroll through the output labelled *mark.out* (you likely want to rename the file) and find the expected values and the “estimates”.

```
modtype-->1     1: 0 0 0 0 0 0 0 0 14.953421
   2: 0 0 0 0 0 0 0 1 0.237180
   3: 0 0 0 0 0 0 1 0 0.237180
   4: 0 0 0 0 0 0 1 1 0.101649
   5: 0 0 0 0 0 1 0 0 0.237180
   6: 0 0 0 0 0 1 0 1 0.101649
   7: 0 0 0 0 0 1 1 0 0.101649
```

Std Design p(*) model

```
========================================================
   Individual Site estimates of <psi>
                Site                 estimate  Std.err   95% conf. interval
psi             1 site_1           :  0.4000   0.1063    0.2187 - 0.6136

==============================================================

   Individual Site estimates of <p1>
                Site                 estimate  Std.err   95% conf. interval
p1              1 site_1           :  0.3000   0.0588    0.1984 - 0.4260
p2              1 site_1           :  0.3000   0.0588    0.1984 - 0.4260
p3              1 site_1           :  0.3000   0.0588    0.1984 - 0.4260
p4              1 site_1           :  0.3000   0.0588    0.1984 - 0.4260
```

Single Species; Single-Season - Planning - EXERCISE II

What is effect of going to $\psi(), p(t)$ model?

StdDesign p(t) model

```
============================================
   Individual Site estimates of <psi>
                    Site                 estimate   Std.err    95% conf. interval
psi                 1 site_1          :   0.4000    0.1063      0.2187 - 0.6136

============================================================

   Individual Site estimates of <p1>
                    Site                 estimate   Std.err    95% conf. interval
p1                  1 site_1          :   0.3000    0.1503      0.0953 - 0.6355
p2                  1 site_1          :   0.3000    0.1503      0.0953 - 0.6355
p3                  1 site_1          :   0.3000    0.1503      0.0953 - 0.6355
p4                  1 site_1          :   0.3000    0.1503      0.0953 - 0.6355
p5                  1 site_1          :   0.3000    0.1503      0.0953 - 0.6355
p6                  1 site_1          :   0.3000    0.1503      0.0953 - 0.6355
p7                  1 site_1          :   0.3000    0.1503      0.0953 - 0.6355
p8                  1 site_1          :   0.3000    0.1503      0.0953 - 0.6355

============================================================
```

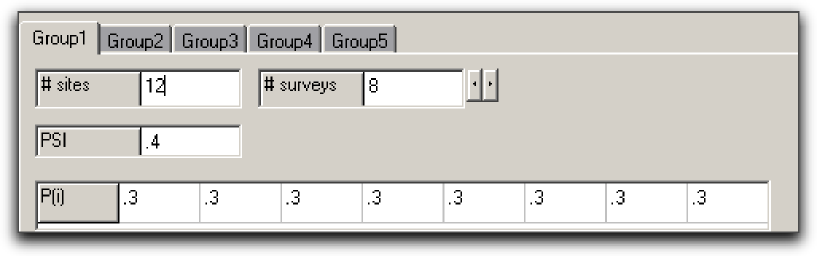
Single Species; Single-Season - Planning - EXERCISE II

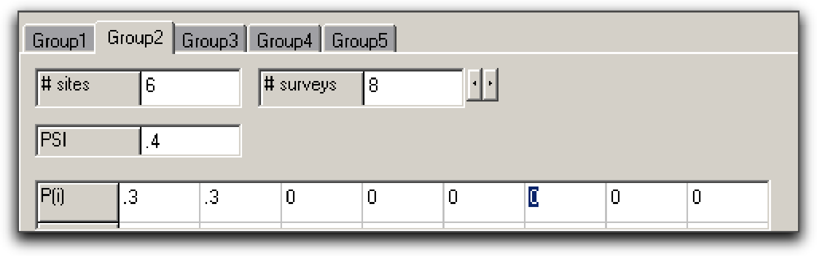
Panel design, 36 sites, 144 surveys

	Survey							
# sites	1	2	3	4	5	6	7	8
12	x	x	x	x	x	x	x	x
6	x	x	-	-	-	-	-	-
6	-	-	x	x	-	-	-	-
6	-	-	-	-	x	x	-	-
6	-	-	-	-	-	-	x	x

Setup 5 groups with $p_s = 0$ if a site not surveyed.
Don't forget to change the number of sites visited.

Single Species; Single-Season - Planning - EXERCISE II





Single Species; Single-Season - Planning - EXERCISE II

Look at expected values to check that doing what you expect:

```
modtype-->1     1:  0  0  0  0  0  0  0  0  7.476710
   2:  0  0  0  0  0  0  0  1  0.118590
   3:  0  0  0  0  0  0  1  0  0.118590
   4:  0  0  0  0  0  0  1  1  0.050824
   5:  0  0  0  0  0  1  0  0  0.118590
   6:  0  0  0  0  0  1  0  1  0.050824
   7:  0  0  0  0  0  1  1  0  0.050824
```

```
257:  0  0  -1  -1  -1  -1  -1  -1  4.776000
258:  0  1  -1  -1  -1  -1  -1  -1  0.504000
259:  1  0  -1  -1  -1  -1  -1  -1  0.504000
260:  1  1  -1  -1  -1  -1  -1  -1  0.216000
261:  -1  -1  0  0  -1  -1  -1  -1  4.776000
262:  -1  -1  0  1  -1  -1  -1  -1  0.504000
263:  -1  -1  1  0  -1  -1  -1  -1  0.504000
```

Single Species; Single-Season - Planning - EXERCISE II

What is impact of Hybrid Design with $p(*)$ model?

Hybrid Design p(*) model

```
========================================================================
   Individual Site estimates of <psi>
                     Site                      estimate   Std.err    95% conf. interval
psi                  1 site_1              :   0.4000     0.1200      0.2002 - 0.6397

============================================================

   Individual Site estimates of <p1>
                     Site                      estimate   Std.err    95% conf. interval
p1                   1 site_1              :   0.3000     0.0757      0.1746 - 0.4648
p2                   1 site_1              :   0.3000     0.0757      0.1746 - 0.4648 |
p3                   1 site_1              :   0.3000     0.0757      0.1746 - 0.4648
p4                   1 site_1              :   0.3000     0.0757      0.1746 - 0.4648
p5                   1 site_1              :   0.3000     0.0757      0.1746 - 0.4648
p6                   1 site_1              :   0.3000     0.0757      0.1746 - 0.4648
p7                   1 site_1              :   0.3000     0.0757      0.1746 - 0.4648
p8                   1 site_1              :   0.3000     0.0757      0.1746 - 0.4648
```

Single Species; Single-Season - Planning - EXERCISE II

What is impact of Hybrid Design with $p(t)$ model?

Hybrid Design p(t) model

```
=============================================================

   Individual Site estimates of <psi>
                     Site                estimate   Std.err    95% conf. interval
psi                  1 site_1          :   0.4000    0.1200     0.2002 - 0.6397

=============================================================

   Individual Site estimates of <p1>
                     Site                estimate   Std.err    95% conf. interval
p1                   1 site_1          :   0.3000    0.1804     0.0737 - 0.6977
p2                   1 site_1          :   0.3000    0.1804     0.0737 - 0.6977
p3                   1 site_1          :   0.3000    0.1804     0.0737 - 0.6977
p4                   1 site_1          :   0.3000    0.1804     0.0737 - 0.6977
p5                   1 site_1          :   0.3000    0.1804     0.0737 - 0.6977
p6                   1 site_1          :   0.3000    0.1804     0.0737 - 0.6977
p7                   1 site_1          :   0.3000    0.1804     0.0737 - 0.6977
p8                   1 site_1          :   0.3000    0.1804     0.0737 - 0.6977
```

Single Species; Single-Season - Planning - EXERCISE II

How many sites would you need to match expected precision from standard design?

Single Species; Single-Season - Heterogeneity

Dealing with (Hidden) Heterogeneity in Detection Probabilities

Single Species; Single-Season - Heterogeneity

Causes of heterogeneity in detection probabilities

- Different habitats, but use habitat then as covariate.
- Different effort, but use models with p_t or covariates.
- (Hidden) different abundance in sites. Sites with higher abundance had higher chance of detecting an "occupation" than sites with lower abundance. Abundance is generally unknown (latent).
- (Hidden) differences among individuals (behaviour, size, home ranges). These cannot be measured
- (Hidden) differences among sites.

Pure Heterogeneity generally causes a negative bias in estimates of occupancy.

Single Species; Single-Season - Heterogeneity

Individual/Site (hidden) heterogeneity - approaches:

- Mixture approach – 2 or more (hidden) groups of animals/sites with different detection rates. Consequences, more histories than expected of the form 0000 or 1111.
- Distribution of detectability approach – animal/site detection rates modeled by distribution. Consequences, more histories than expected of the form 0000 or 1111.

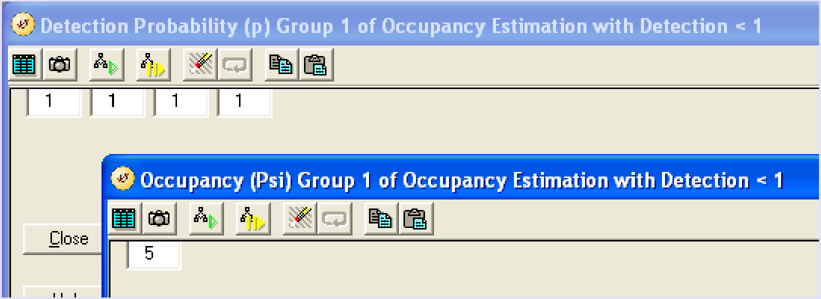
In both cases, the group members is NOT known (latent).

Single Species; Single-Season - Heterogeneity - *MARK*

Finite Mixture Approach - Bull Frog Example

500 ponds were visited at 4 occasions and the observer spent 30 minutes listening for call.

Open MARK, and import the *bullfrog.inp* data in the usual fashion. Fit the $\psi(*)$, $p(*)$ model.



Single Species; Single-Season - Finite Mixture Models - *MARK*

$\psi(*), p(*)$ model results

Model	AICc	Delta AICc	AICc Weight	Model Likelihood	No. Par.	Deviance
{psi(*) p(*)}	1891.2920	0.0000	1.00000	1.0000	2	33.0193

Bull Frogs

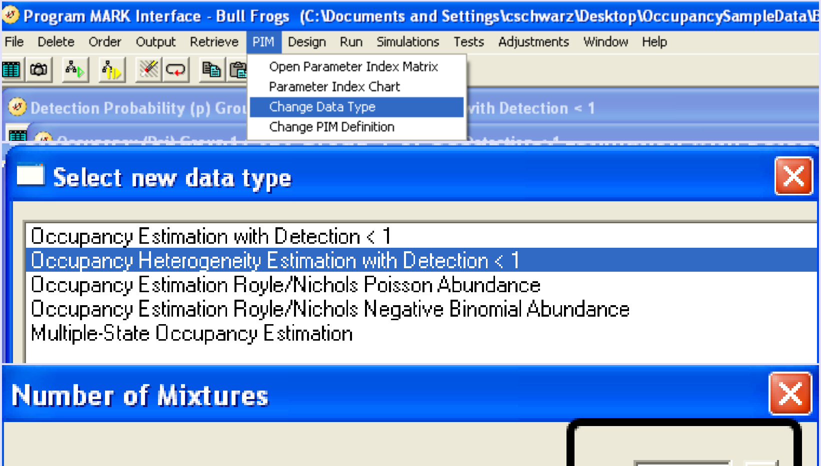
Real Function Parameters of {psi(*) p(*)}

Parameter	Estimate	Standard Error	95% Confidence Interval Lower	Upper
1:p	0.5519311	0.0183166	0.5158217	0.5875012
2:Psi	0.4647319	0.0233824	0.4193258	0.5107302

But the Bootstrap GOF gives evidence of a poor fit (goodness-of-fit p-value < 0.01).

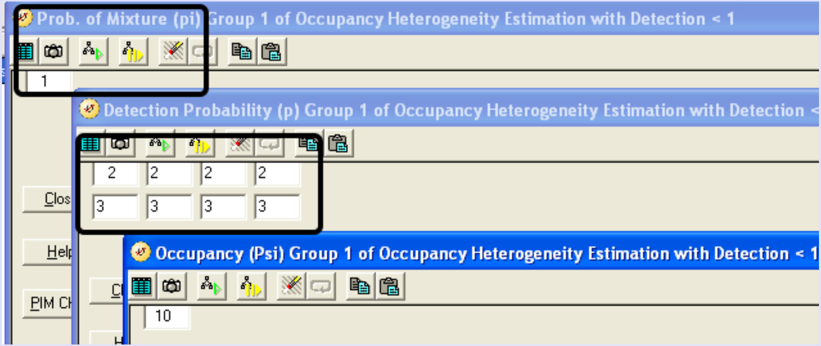
Single Species; Single-Season - Finite Mixture Models - *MARK*

Change the data type to a Finite-Mixture Model (usually 2 groups is sufficient):



Single Species; Single-Season - Finite Mixture Models - *MARK*

There is a mixing proportion (π) (why only 1 value?) and detection has two rows corresponding to the two (latent) groups.



Run this model $\psi(*)$, p(* - 2 groups).

May need to try different initial values if $\widehat{\psi}$ wanders off to 1 or 0.

Single Species; Single-Season - Finite Mixture Models - *MARK*

Model $\psi(*)$, p(* - 2 groups)

Results Browser: Occupancy Heterogeneity Estimation with Detection < 1

Model	AICc	Delta AICc	AICc Weight	Model Likelihood	No. Par.	Deviance	-2Log(L)
{psi(*) p(* - 2 groups)}	1870.8463	0.0000	0.99996	1.0000	4	8.5169	1862.7655
{psi(*) p(*)}	1891.2920	20.4457	0.00004	0.0000	2	33.0193	1887.2679

Bull Frogs

Real Function Parameters of {psi(*) p(* - 2 groups)}

Parameter	Estimate	Standard Error	95% Confidence Interval Lower	95% Confidence Interval Upper
1:pi	0.7711791	0.1173249	0.4779714	0.9254033
2:p	0.4205930	0.0614628	0.3068949	0.5433898
3:p	0.8772993	0.0956976	0.5559257	0.9760970
4:Psi	0.4884813	0.0280898	0.4337851	0.5434546

Deviance is much improved (unfortunately Bootstrap GOF not yet available). ΔAIC is large relative to other model.

Careful not to OVER INTERPRET the results – groups is just a DEVICE to account for heterogeneity – they often have no physical interpretation.

Single Species; Single-Season - Royle-Nichols models

Suppose that there are N_i individuals in a site, then the probability of a detection is

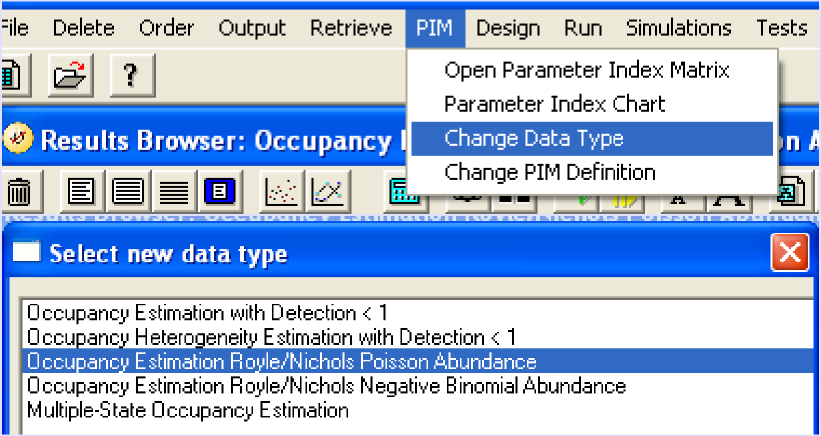
$$p(N_i, r) = 1 - (1 - r)^{N_i}$$

where r is the individual detection probability. Note that N_i is not observed, but if you are willing to assume that N_i comes from some distribution (e.g. Poisson with parameter λ (average density/site), you can "integrate" over the hidden N_i.
Now $\psi = 1 - e^{-\lambda}$. Notice that ψ is a DERIVED parameter and does NOT appear in the model directly.

CAUTION: This class of models is VERY sensitive to assumptions about the distribution of the N_i.

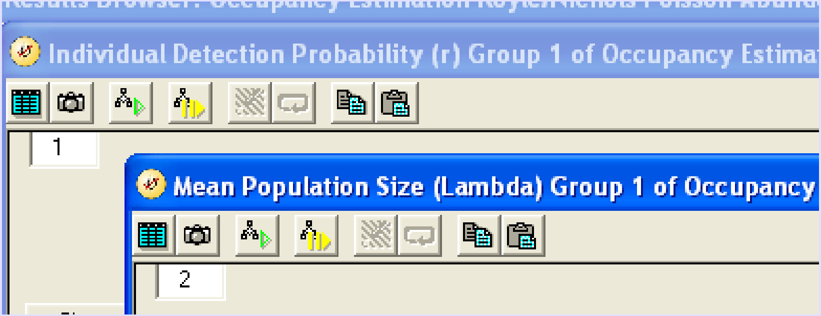
Single Species; Single-Season - Royle-Nichols models

Change data type



Single Species; Single-Season - Royle-Nichols models

Royle-Nichols Poisson model: PIM is very simple! Notice that there is NO PIM for ψ as it is a derived parameter.



Single Species; Single-Season - Royle-Nichols models

{psi(*) p(* - 2 groups)}	1870.8463	0.0000	0.86395	1.0000	4	8.5169	1862.7655
{Royle-Nichols Poisson Model}	1874.5438	3.6975	0.13601	0.1574	2	16.2710	1870.5197
{psi(*) p(*)}	1891.2920	20.4457	0.00003	0.0000	2	33.0193	1887.2679
{psi(*) p(t)}	1893.9806	23.1343	0.00001	0.0000	5	29.6104	1883.8591

```
                                 Bull Frogs

              Real Function Parameters of {Royle-Nichols Poisson Model}
                                                       95% Confidence Interval
 Parameter                    Estimate       Standard Error      Lower           Upper
 --------------------------  --------------  --------------  --------------  --------------
    1:r                       0.4561290       0.0230632       0.4114064       0.5015705
    2:Lambda                  0.6591343       0.0483081       0.5592245       0.7466571
```

```
                       Bull Frogs

              Estimates of Derived Parameters
       Occupancy Estimates of {Royle-Nichols Poisson Model}
                                       95% Confidence Interval
 Group   Psi-hat        Standard Error      Lower          Upper
 -----  --------------  --------------  --------------  --------------
   1    0.4827010       0.0249897       0.4337212       0.5316809
       E(p-hat) Estimates of {Royle-Nichols Poisson Model}
                                       95% Confidence Interval
 Group   E(p-hat)       Standard Error      Lower          Upper
 -----  --------------  --------------  --------------  --------------
   1    0.5379383       0.0211306       0.4965222       0.5793544
```

Single Species; Single-Season - Heterogeneity Summary

Summary about heterogeneity in catchability:

- Usually leads to underestimation of ψ.
- Largest effects when $p \approx 0$!
- Mixture-models require substantial data to work well.
- Better to design studies to minimize heterogeneity.
- Measure relevant covariates.
- If detectability related to abundance and abundance varies considerably, occupancy modeling not my first choice as estimates are (highly) sensitive to assumptions and minor changes in the data.

Single Species; Single-Season - Summary

Single-Species Single-Season Summary

Single Species; Single-Season - Summary

Planning.

- Key parameter ψ; nuisance parameter $p_t < 1$.
- Design your study well.
  - What is appropriate spatial scale?
  - Simple Random sample of sites; some relaxation if comparing occupancy between classes.
  - What is a season – assume closure over season.
  - Repeated surveys must be independent.
- Allocate effort between sites and surveys; usually more sites and fewer surveys (but at least 2).
- Hybrid designs may be more efficient. CAUTION about "removal" design. Consider panel designs.

Single Species; Single-Season - Summary

Key assumptions.

1. Occupancy state of sites is constant during all single-season surveys (closure).
2. Probability of occupancy (ψ) is equal across all sites (homogeneity).
3. Probability of detection (p) given occupancy is equal across all sites (homogeneity).
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
6. No false positives.

Some models available that deal with violations of assumptions, but these are data hungry.

Single Species; Single-Season - Summary

Analysis

1. Maximum Likelihood & AIC & Model average
2. Software
    - *MARK*, *RMark*
    - *PRESENCE*, *RPresence*
    - *R* package *unmarked*
    - *JAGS* Bayesian models.
    - GENPRES for planning
3. Carefully think of models and biological realism.
4. Do not data dredge.

Garbage in $\rightarrow$ garbage out.