Design and Analysis of Occupancy Studies
Part 1a

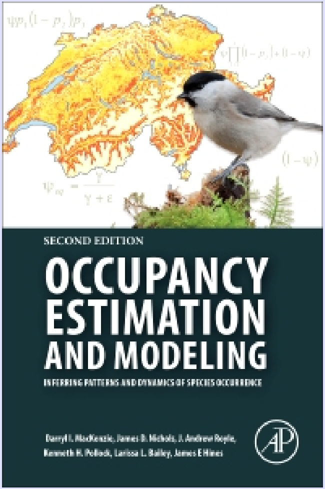
Carl James Schwarz

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

Outline I

Outline II

Books on Occupancy Modeling



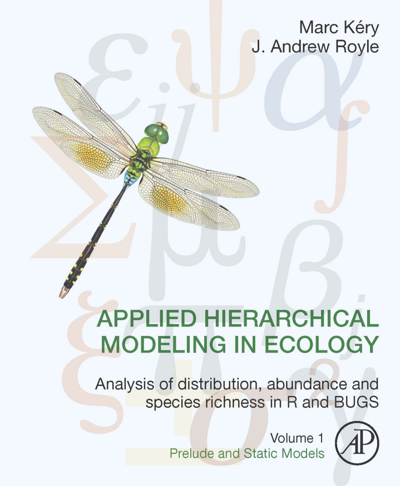
SECOND EDITION

OCCUPANCY
ESTIMATION
AND MODELING

INFERRING PATTERNS AND DYNAMICS OF SPECIES OCCURRENCE

Darryl I. MacKenzie, James D. Nichols, J. Andrew Royle,
Kenneth H. Pollock, Larissa L. Bailey, James E Hines

Books on Occupancy Modeling



Marc Kéry
J. Andrew Royle

APPLIED HIERARCHICAL MODELING IN ECOLOGY

Analysis of distribution, abundance and species richness in R and BUGS

Volume 1
Prelude and Static Models

What is Occupancy Modeling

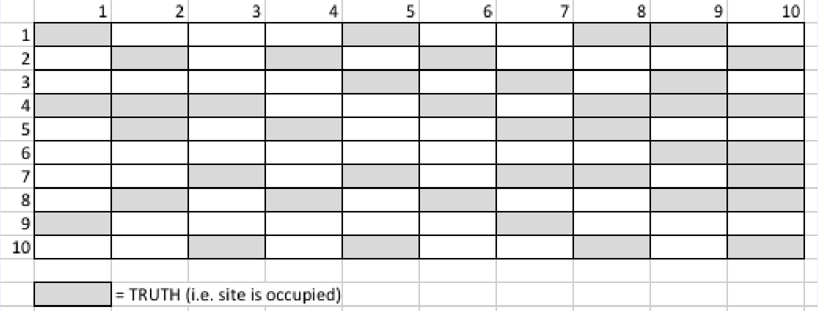
Occupancy: Proportion of area/ patches/ sample units that are occupied by a species of interest.

- Studies of distribution and range, e.g. spotted owl or bull trout
- Is patch occupancy related to patch characteristics (e.g. habitat variables)?
- Spread of invasive species.
- Disease dynamics.
- Pattern of species richness in space and time
- Species assemblages – which species tend to co-occur or anti-occur?

Key problem: False negatives, i.e. a survey may not detect an animal present on patch because animal is low density, cryptic, observer error, etc. **Detection probability** < 1**!**
Naive estimates lead to underestimates of occupancy, species numbers, etc.

Basic idea of occupancy studies

Landscape is divided into patches or cells (not necessary that they be equal size), some of which are occupied by species of interest.



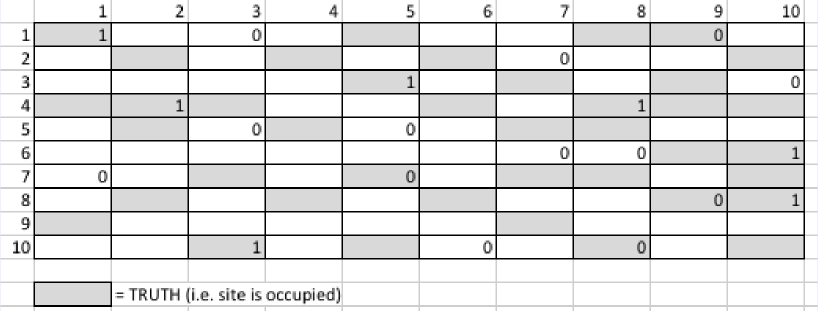
	1	2	3	4	5	6	7	8	9	10
1	▨				▨			▨	▨	
2		▨		▨		▨				▨
3					▨		▨		▨	
4	▨	▨	▨			▨		▨	▨	▨
5		▨		▨			▨	▨		
6									▨	▨
7			▨		▨		▨	▨		▨
8		▨		▨		▨			▨	▨
9	▨						▨			
10			▨		▨			▨		▨

▨ = TRUTH (i.e. site is occupied)

$\psi = 0.40$ (TRUTH $=$ proportion of sites occupied in population)

Basic idea of occupancy studies

Suppose we sample 20 cells (at random) with a detection probability < 1 with a single visit/cell.



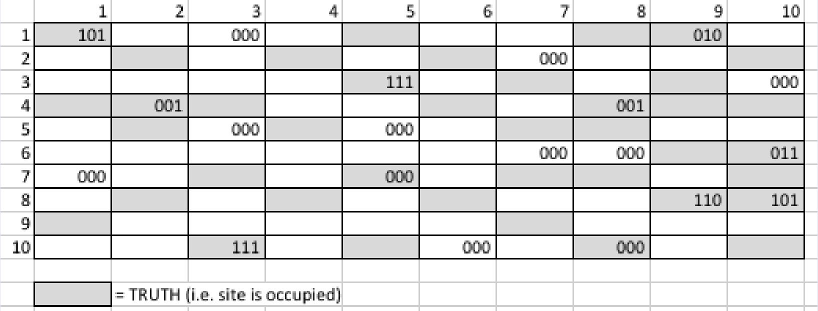
	1	2	3	4	5	6	7	8	9	10
1	▨ 1		0		▨			▨	▨ 0	
2		▨		▨		▨	0			▨
3					▨ 1		▨		▨	0
4	▨	▨ 1	▨			▨		▨ 1	▨	▨
5		▨	0	▨	0		▨	▨		
6							0	0	▨	▨ 1
7	0		▨		▨ 0		▨	▨		▨
8		▨		▨		▨			▨ 0	▨ 1
9	▨						▨			
10			▨ 1		▨	0		▨ 0		▨

▨ = TRUTH (i.e. site is occupied)

Notice the false negatives, e.g. in row 10 and column 8.
$\psi = 0.40$ (TRUTH = proportion of sites occupied in population)
$\widehat{\psi}_{naive} = \frac{7}{20} = 0.35$ UNDERESTIMATES true occupancy because of false negatives.

Basic idea of occupancy studies

Suppose we sample 20 cells (at random) with detection probability < 1 with 3 visits/cell.



	1	2	3	4	5	6	7	8	9	10
1	▨ 101		000		▨			▨	▨ 010	
2		▨		▨		▨	000			▨
3					▨ 111		▨		▨	000
4	▨	▨ 001	▨			▨		▨ 001	▨	▨
5		▨	000	▨	000		▨	▨		
6							000	000	▨	▨ 011
7	000		▨		▨ 000		▨	▨		▨
8		▨		▨		▨			▨ 110	▨ 101
9	▨						▨			
10			▨ 111		▨	000		▨ 000		▨

▨ = TRUTH (i.e. site is occupied)

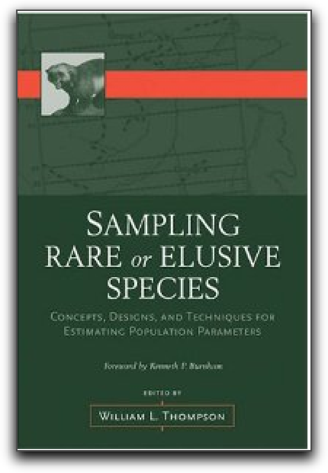
$\psi = 0.40$ (TRUTH = proportion of sites occupied in population)
Empty cells are always 0; Occupied cells with history such as 101 provide information about the detection rate because cell was occupied but not detected on second visit.

When to use Occupancy Modeling

- Interest at SPECIES level and not individual level.
- Discrete patches (SITES) where species are relatively immobile (e.g. ponds and frogs; forest stands and invasive plants).
- Detection rates < 1 but not too small (i.e. around 0.3 and higher).
- Occupancy between 0.2 and 0.8.

When to NOT use Occupancy Modeling

Very rare or elusive species. See

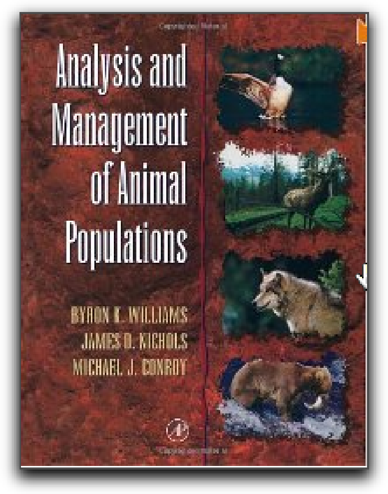


Thompson, W. (2004). Sampling Rare or Elusive Species: Concepts, Designs, and Techniques for Estimating Population Parameters.

Problem is that most of encounter histories will be 000000 and so very difficult to estimate detection rate and extra false-negatives.

When to NOT use Occupancy Modeling

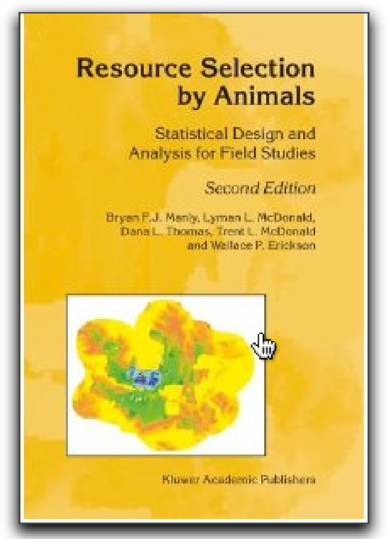
Interested in population parameters such as abundance, survival, movement, etc.



Williams, B.K, Nichols, J.D. and Conroy, M.J. (2002).
Analysis and Management of Animal Populations.

When to NOT use Occupancy Modeling

Resource selection.



Resource Selection by Animals

Statistical Design and Analysis for Field Studies

Second Edition

Bryan F.J. Manly, Lyman L. McDonald, Dana L. Thomas, Trent L. McDonald and Wallace P. Erickson

Kluwer Academic Publishers

Manly, B.F., McDonald, L., Thomas, D. L., McDonald,T.L. and Erickson, W.P. (2002). Resource Selection by Animals: Statistical Design and Analysis for Field Studies.

Types of occupancy studies

Single species; single season

- Sampling protocol:
    - Multiple visits to random sample of sites;
    - Record OBSERVED detections/non-detections.
- Estimate
    - Proportion of sites that are occupied (ψ)
    - Detection probability p on each visit.
    - Variation in parameters as functions of covariates such as patch size, habitat, time, management actions, etc.

Types of occupancy studies

Single species; multiple seasons

- Sampling protocol:
  - Multiple visits to random sample of sites;
  - Record OBSERVED detection/non-detections;
  - Repeat over multiple seasons
- Estimate
  - Proportion of sites that are occupied (ψ)
  - Detection probability p on each visit.
  - Local extinction and colonization probabilities
  - Variation in parameters as functions of covariates such as patch size, habitat, time, management actions, etc.

Types of occupancy studies

Multiple species; single and multiple seasons

- Sampling protocol:
  - Multiple visits to random sample of sites;
  - Record OBSERVED detection/non-detection of EACH species
  - Single or multiple seasons
- Estimate
  - Proportion of sites that are occupied (ψ) for each species
  - Dynamics - local colonization and extinction probabilities by species and relative to other species
  - Detection probability p on each visit.
  - Variation in parameters as functions of covariates such as patch size, habitat, time, management actions, etc.

Types of occupancy studies

Multiple state; single and multiple seasons

- Sampling protocol:
    - Multiple visits to random sample of sites;
    - Record OBSERVED state of species (e.g. not detected, detected adults, detected breeders)
    - Single or multiple seasons
- Estimate
    - Proportion of sites that are occupied (ψ).
    - What fraction of occupied sites are in each STATE (e.g. breeders and non-breeders).
    - Detection probability p for each visit.
    - Local extinction and colonization probabilities.
    - Variation in parameters as functions of covariates such as patch size, habitat, time, management actions, etc.

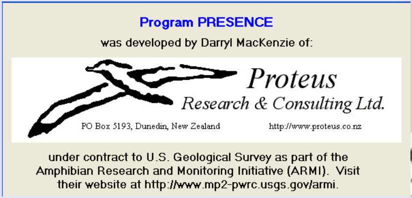
Types of occupancy studies

Multiple scale; single and multiple seasons

- Sampling protocol:
    - Multiple visits to random sample of sites;
    - Record OBSERVED detection/non-detection using different methods (e.g. visual or aural) or at different scales (small quadrats or larger landscape).
    - Single or multiple seasons
- Estimate
    - Proportion of sites that are occupied (ψ).
    - Detection probability p by methods/scale.
    - Local extinction and colonization probabilities.
    - Variation in parameters as functions of covariates such as patch size, habitat, time, management actions, etc.

Software for Occupancy Modeling

- PRESENCE
- GENPRES (planning studies)
- MARK
- *R* package *RPresence*
- *R* package *RMark*
- *R* package *unmarked* (not recommended)
- Bayesian modelling using *JAGS* for complex problems.



Program PRESENCE

was developed by Darryl MacKenzie of:

Proteus
Research & Consulting Ltd.
PO Box 5193, Dunedin, New Zealand http://www.proteus.co.nz

under contract to U.S. Geological Survey as part of the Amphibian Research and Monitoring Initiative (ARMI). Visit their website at http://www.mp2-pwrc.usgs.gov/armi.

Program MARK

You can obtain context-sensitive help with the F1 key, and can investigate objects with the Shift-F1 key. See the Help menu for known problems.

Single Species; Single-Season - Sampling Protocol

Single-Species Single-Season Occupancy Studies

Theory

Single-Species Single-Season - Sampling Protocl

Sampling Protocol:

- Landscape divided (artificially or naturally) into S patches or cells or SITES.
- Select $s << S$ sites at random (all sites have equal probability of selection)
  - Randomization makes sample representative.
  - Sample size controls precision.
  - Stratification (see later in course) may be useful
- Visit each site K times (unequal visits covered later)
- Record detection or not detection of species in site i in each visit k.
  - Visual aural, fecal, browse, signs, etc
- Create a **Detection/Encounter History** for each surveyed site (e.g. 0110).

Single Species; Single-Season - 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.

Single Species; Single-Season - Theory

Important terms:

- **Population**: Complete set of sites for which inference is wanted
- **Sample**: Sites actually measured
- **Parameter**: Occupancy over entire population (ψ) - always unknown
- **Estimate**: Values obtained from the sample data ($\widehat{\psi}$).

The RRR's of statistics:

- **Randomization**: Makes a sample representative.
- **Replication**: Control precision (the SE)
- **Stratification**: Account for one form of heterogeneity in population.

Single Species; Single-Season - Theory

Every SAMPLE will give a DIFFERENT estimate!

A measure of how much the estimate could vary if a new sample is taken is the STANDARD ERROR (SE).

A usual convention is to report a 95% CONFIDENCE INTERVAL which is a plausible range for the POPULATION parameter given the collected data. A c.i. does NOT make statement about individual values (i.e. the occupancy of a particular site).

Single Species; Single-Season - Theory - MLE

LIKELIHOOD connects the data (detect/not detect) with the parameters of the model (ψ and p). Need to construct a probability expression for each history in the sample. There are two cases:

- **At least one detection**. Consider the history (10100). Then

$$P(10100) = \psi \times p(1-p)p(1-p)(1-p)$$

- **No detections**. Either occupied and not seen or not occupied:

$$P(00000) = \psi(1-p)^5 + (1-\psi)$$

Single Species; Single-Season - Theory - MLE

Likelihood function:

$$L = P(h_1) \times P(h_2) \times \ldots \times P(h_s) = L(data, parameters)$$

Find the values of the parameter (the MLEs) that MAXIMIZE the overall probability of the data. In simple cases this can be done by hand, but usually is done numerically.

The SE can also be obtained from derivatives of the (log)-likelihood function

Single Species; Single-Season - Theory - MLE

Properties of MLEs

- Extracts ALL of the information from the data
- Give the smallest possible SE
- Are unbiased (in large samples)
- Enable CI to be constructed as $Estimate \pm 2SE$.

However, in small samples, MLEs are not guaranteed to be optimal.

Single Species; Single-Season - A new model

We assumed that p was equal across all sampling occasions. But suppose that effort/ conditions varied across occasions. Need to relax the assumption that p was equal across all occasions.

Standard Notation:

- Model $\psi(*), p(*)$ had equal p across all ocasions.
- Model $\psi(*), p(t)$ has DIFFERENT p for each occasion.

Write out P(history) for history 10100 00001 and 00000.

Single Species; Single-Season - AIC

Which model is best?

"All models are wrong, but some are useful" (George Box)

AIC = Akaike Information Criteria = tradeoff between fit and complexity

$$AIC = -2\log likelihood + 2\#p$$

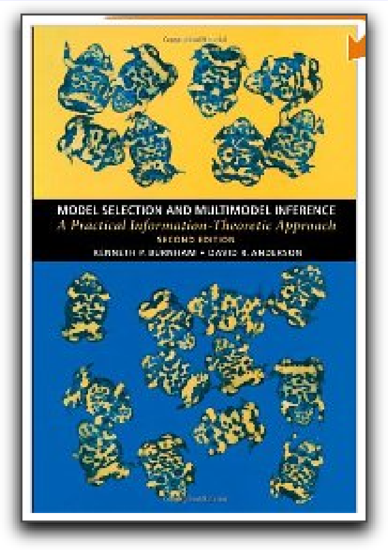
where $\#p$ is the number of parameters.

"Best" model in the set is one with arithmetically smallest AIC. Find

$$\Delta AIC_m = AIC_m - AIC_{min}$$

Differences of 2 or 3 AIC units imply no real preference.

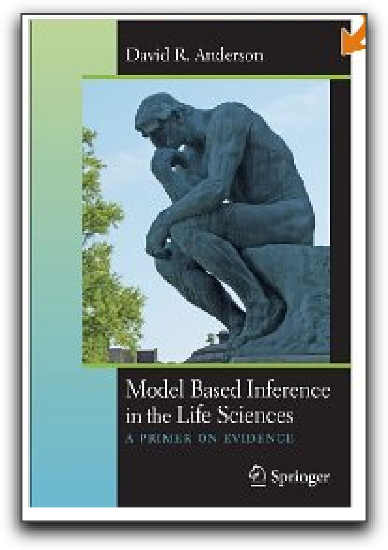
Single Species; Single-Season - AIC



Burnham, K.P. and Anderson, D.R. (2002).
Model selection and Multi-Model Inference: A practical information-theoretic approach.

Complete but dense. The first few chapters are very readable.

Single Species; Single-Season - AIC



Anderson, D.R. (2008).
Model based inference in the life sciences: A primer on evidence.

Very readable explanation of the AIC paradigm.

Single Species; Single-Season - AIC

Model averaging - compute a model weight

$$w_m = \frac{\exp(-\frac{\Delta AIC_m}{2})}{\sum \exp(-\frac{\Delta AIC_i}{2})}$$

Interpret as the weight of evidence in favor of model m relative to OTHER MODELS in the model set.

Add up weights for models with a common feature, e.g. habitat variable (see later in notes).

Single Species; Single-Season - AIC

Model averaging - average estimates

$$\widehat{\psi}_{average} = \sum w_m \widehat{\psi}_m$$

This accounts for both WITHIN and AMONG model uncertainty.

Also able to obtain SE that are adjusted for model averaging (weighted average of the individual standard errors and a penalty if the estimates are not consistent among models).

Single Species; Single-Season - Presence

Single-Species Single-Season Occupancy Studies

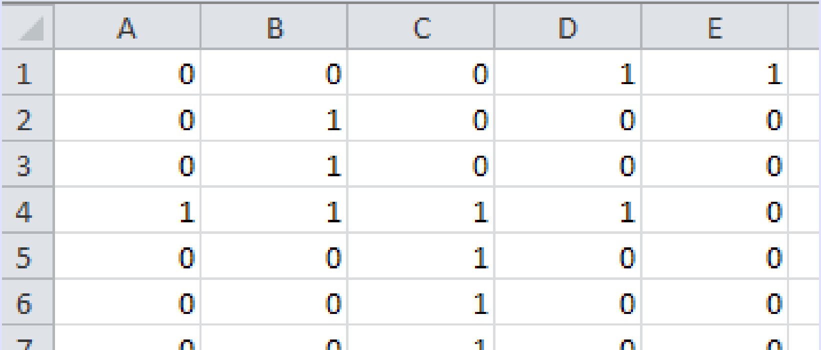
Using Presence

Single Species; Single-Season - PRESENCE - Salamander

Blue-ridge two-lined salamander (*Eurycea wilderae*) occurs in Great Smokey Mountains National Park. 39 sites surveyed every 2 weeks. At each visit, a transect was searched every 10 m by overturning rocks or cover boards.

Open the **salamader.xls** file in *OccupancySampleData* folder

What is naive estimate of occupancy?

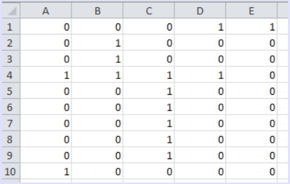


	A	B	C	D	E
1	0	0	0	1	1
2	0	1	0	0	0
3	0	1	0	0	0
4	1	1	1	1	0
5	0	0	1	0	0
6	0	0	1	0	0
7	0	0	1	0	0

Single Species; Single-Season - PRESENCE - Salamander

Create an Excel file with the history for this species in this season
The values in the columns have

- 0 - not detected
- 1 - detected
- . - not visited



	A	B	C	D	E
1	0	0	0	1	1
2	0	1	0	0	0
3	0	1	0	0	0
4	1	1	1	1	0
5	0	0	1	0	0
6	0	0	1	0	0
7	0	0	1	0	0
8	0	0	1	0	0
9	0	0	1	0	0
10	1	0	0	0	0

Single Species; Single-Season - PRESENCE - Salamander

Open PRESENCE

Program PRESENCE 2.12.8

Start a new analysis by clicking File/New Project

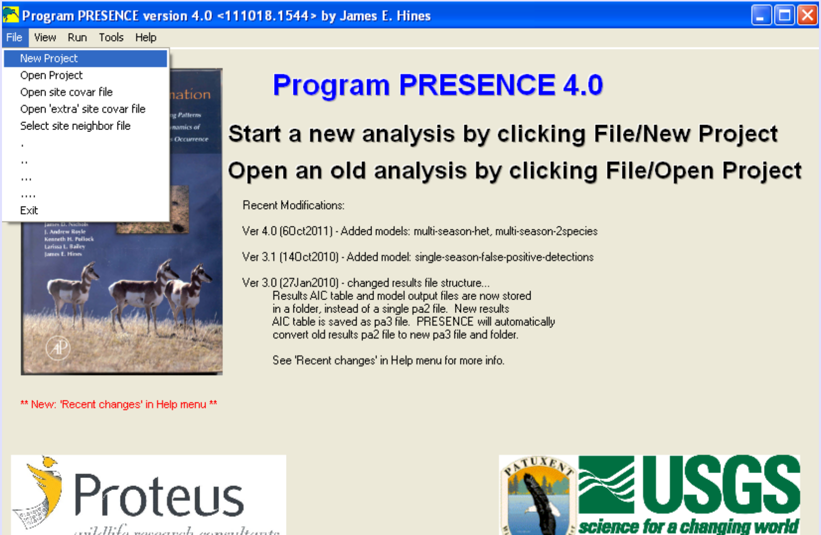
Open an old analysis by clicking File/Open Project

Recent Modifications:
Ver 12.8 (08Aug2018) - added multi-season false-positive model to RPresence
Ver 12.6 (08Aug2018) - no change - update ver number to match RPresence

** New: 'Recent changes' in Help menu **

Single Species; Single-Season - PRESENCE - Salamander

Start a New Project



Program PRESENCE 4.0

Start a new analysis by clicking File/New Project

Open an old analysis by clicking File/Open Project

Recent Modifications:

Ver 4.0 (6Oct2011) - Added models: multi-season-het, multi-season-2species

Ver 3.1 (14Oct2010) - Added model: single-season-false-positive-detections

Ver 3.0 (27Jan2010) - changed results file structure...
Results AIC table and model output files are now stored in a folder, instead of a single pa2 file. New results AIC table is saved as pa3 file. PRESENCE will automatically convert old results pa2 file to new pa3 file and folder.

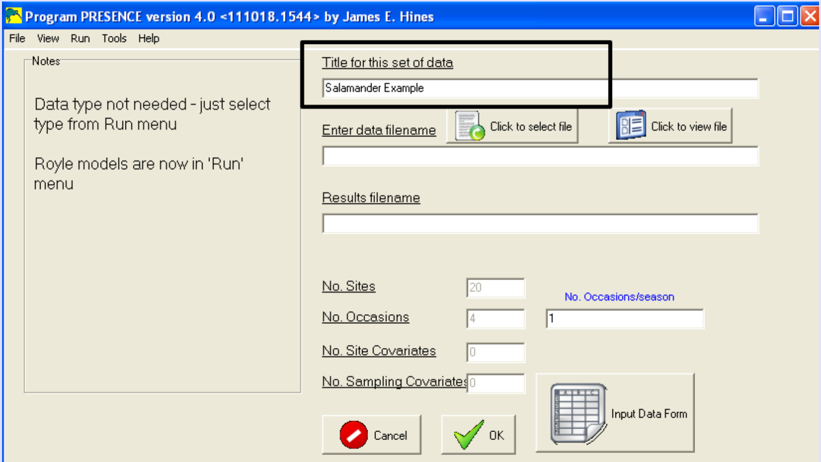
See 'Recent changes' in Help menu for more info.

** New: 'Recent changes' in Help menu **

Single Species; Single-Season - PRESENCE - Salamander

Getting data into PRESENCE is a bit tricky as the work flow is not top down.

Enter a title for the project.



Program PRESENCE version 4.0 <111018.1544> by James E. Hines

Notes

Data type not needed - just select type from Run menu

Royle models are now in 'Run' menu

Title for this set of data: Salamander Example

Enter data filename:

Results filename:

No. Sites: 20

No. Occasions: 4

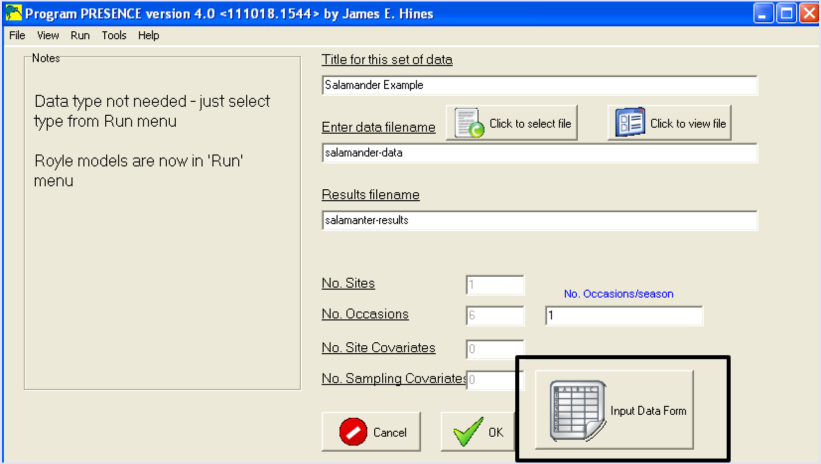
No. Occasions/season: 1

No. Site Covariates: 0

No. Sampling Covariates: 0

Single Species; Single-Season - PRESENCE - Salamander

Import the data by clicking on *Input Data Form*



Notes

Data type not needed - just select type from Run menu

Royle models are now in 'Run' menu

Title for this set of data: Salamander Example

Enter data filename: salamander-data

Results filename: salamanter-results

No. Sites: 1

No. Occasions: 6

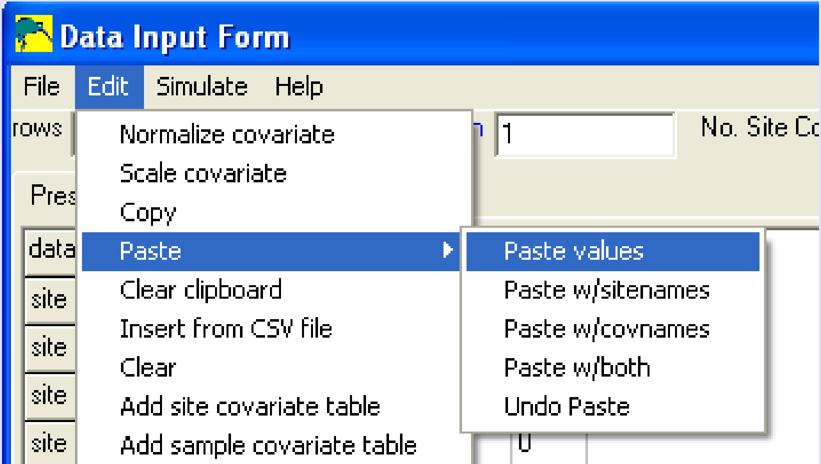
No. Occasions/season: 1

No. Site Covariates: 0

No. Sampling Covariates: 0

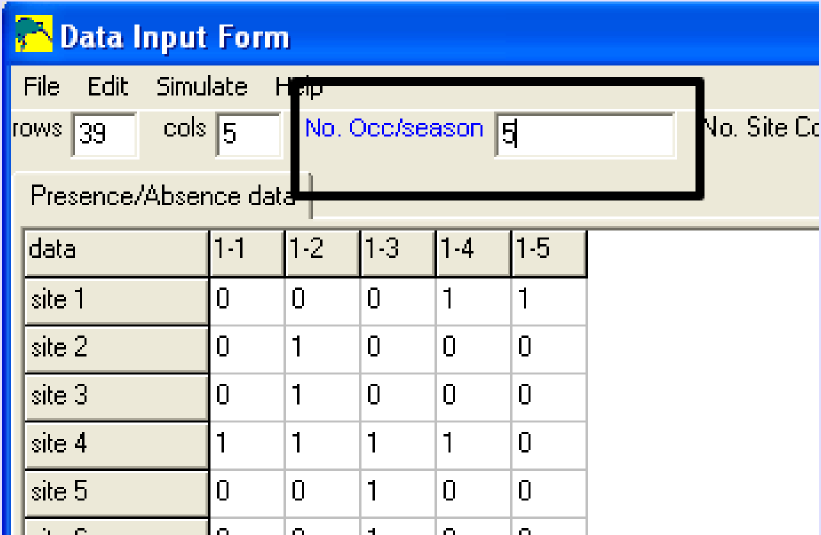
Single Species; Single-Season - PRESENCE - Salamander

Copy/paste the data from the **salamander.xls** to the data entry window. Note the various options to allow you to enter the site name in the first column, the date in the first row, etc.



Single Species; Single-Season - PRESENCE - Salamander

Update the number of occasions per season

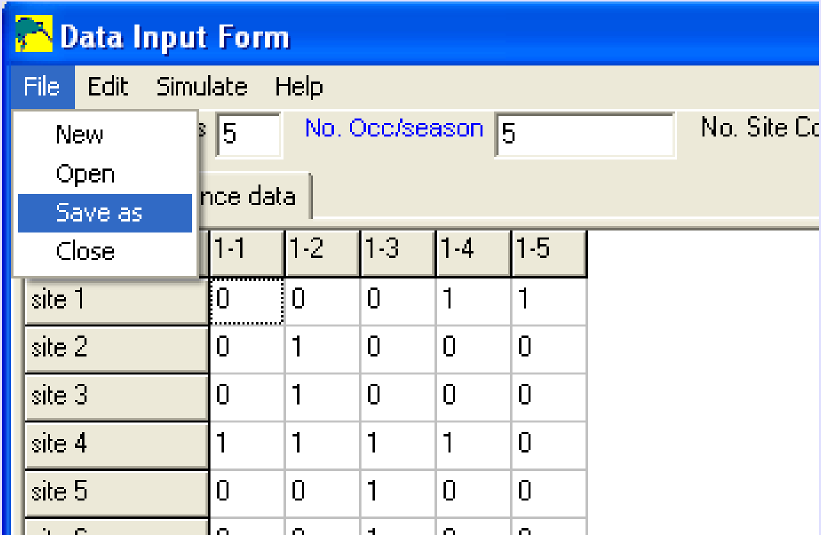


rows 39 cols 5 No. Occ/season 5 No. Site Co

data	1-1	1-2	1-3	1-4	1-5
site 1	0	0	0	1	1
site 2	0	1	0	0	0
site 3	0	1	0	0	0
site 4	1	1	1	1	0
site 5	0	0	1	0	0

Single Species; Single-Season - PRESENCE - Salamander

Save the history + additional information as a *.pao file.



s 5

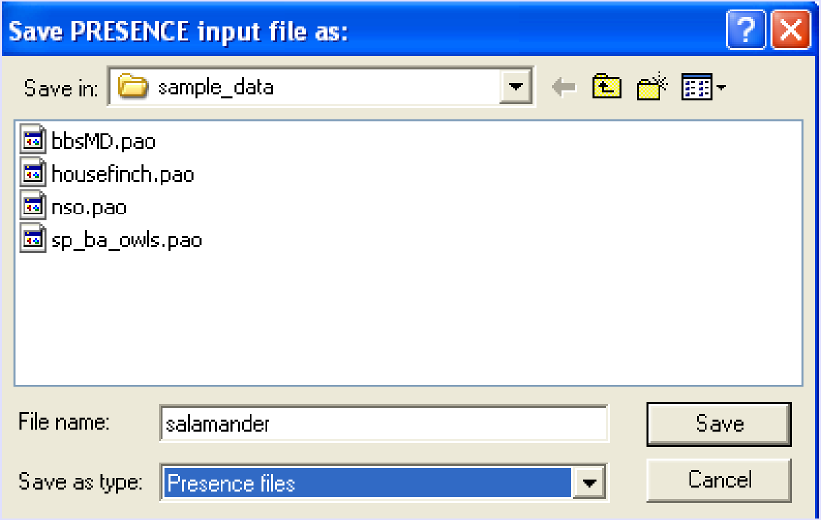
No. Occ/season 5

No. Site Co

	1-1	1-2	1-3	1-4	1-5
site 1	0	0	0	1	1
site 2	0	1	0	0	0
site 3	0	1	0	0	0
site 4	1	1	1	1	0
site 5	0	0	1	0	0

Single Species; Single-Season - PRESENCE - Salamander

Save the history + additional information as a *.pao file.



Save PRESENCE input file as:

Save in: sample_data

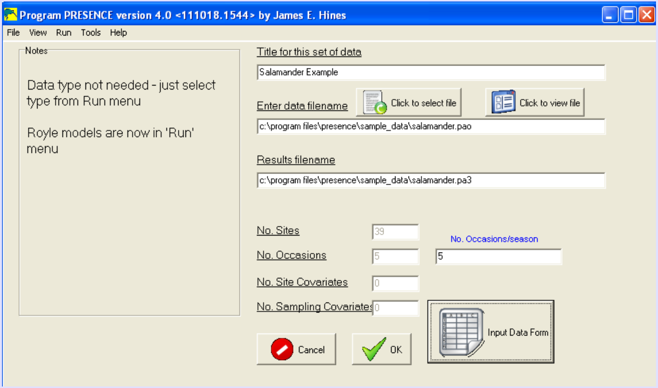
bbsMD.pao
housefinch.pao
nso.pao
sp_ba_owls.pao

File name: salamander

Save as type: Presence files

Single Species; Single-Season - PRESENCE - Salamander

This updates the initial project screen. CHECK INPUT!



Program PRESENCE version 4.0 <111018.1544> by James E. Hines

Notes

Data type not needed - just select type from Run menu

Royle models are now in 'Run' menu

Title for this set of data: Salamander Example

Enter data filename: c:\program files\presence\sample_data\salamander.pao

Results filename: c:\program files\presence\sample_data\salamander.pa3

No. Sites: 39

No. Occasions: 5

No. Occasions/season: 5

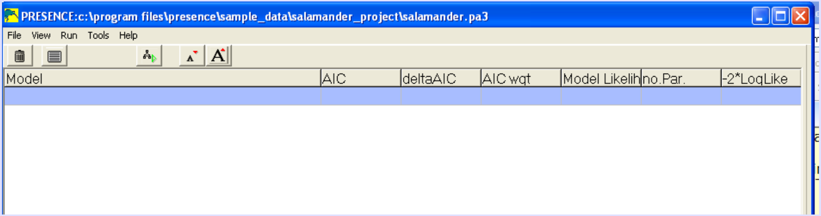
No. Site Covariates: 0

No. Sampling Covariates: 0

Press the check button to create the folder with results, the *pao (data + other information file), the *pa3 (model results file), and the directory.

Single Species; Single-Season - PRESENCE - Salamander

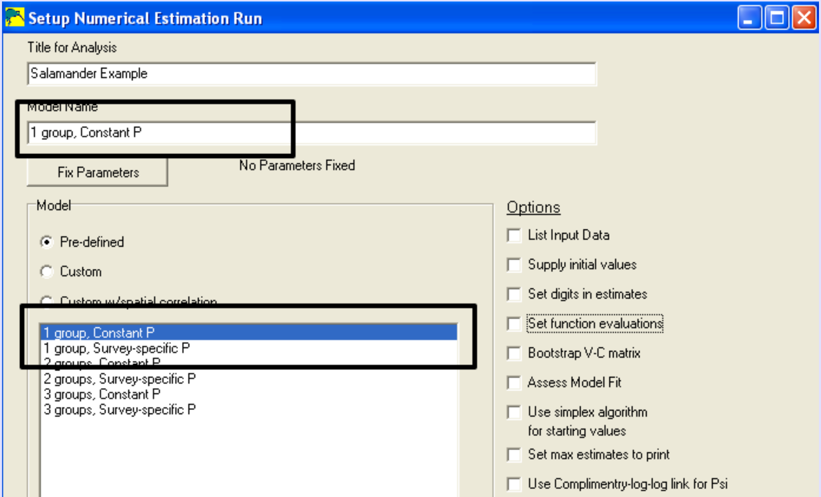
You are now at the RESULTS window.



Explore the RESULTS Window menu items. Notice the types of models that can be fit.

Single Species; Single-Season - PRESENCE - Salamander

Select the Single Season/Single Species and fit the simplest model $(\psi(*), p(*))$ (Notation will become clearer later).



Title for Analysis: Salamander Example

Model Name: 1 group, Constant P

No Parameters Fixed

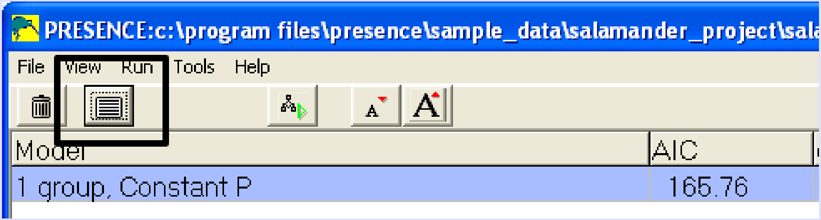
- ● Pre-defined
- ○ Custom
- ○ Custom w/spatial correlation

1 group, Constant P
1 group, Survey-specific P
2 groups, Constant P
2 groups, Survey-specific P
3 groups, Constant P
3 groups, Survey-specific P

- ☐ List Input Data
- ☐ Supply initial values
- ☐ Set digits in estimates
- ☐ Set function evaluations
- ☐ Bootstrap V-C matrix
- ☐ Assess Model Fit
- ☐ Use simplex algorithm for starting values
- ☐ Set max estimates to print
- ☐ Use Complimentry-log-log link for Psi

Single Species; Single-Season - PRESENCE - Salamander

Open the results for this model.

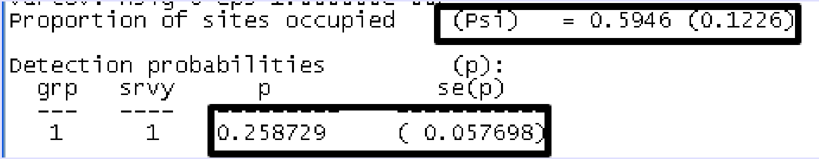


Lot of interesting stuff!

- Basic statistics about your data - make sure these are correct
- Information about fitting the model - typically not of interest.
- *Beta* estimates - typically not of interest
- Estimates of occupancy and detectability.
- Posterior probability of occupancy.

Single Species; Single-Season - PRESENCE - Salamander

Look at output. In particular the estimates:



```
Proportion of sites occupied    (Psi)   = 0.5946 (0.1226)

Detection probabilities                 (p):
  grp    srvy        p                se(p)
  ---    ----     --------        ----------
   1       1    0.258729        ( 0.057698)
```

We have $\widehat{\psi} = 0.59$ (SE 0.12) and $\widehat{p} = 0.26$ (SE 0.06).
Naive estimate of occupancy of 0.46.

Single Species; Single-Season - PRESENCE - Salamander

Probability of missing a species if present over all 5 occasions (probability of a false negative) is found as

$$\widehat{P}(False\ negative \mid occupied) = (1-p)^5 = (1-0.26)^5 = 0.22$$

Conditional probability that a site with no animals observed is occupied is

$$\begin{aligned}\widehat{P}(occupied \mid not\ detected) =& \frac{\widehat{P}(occupied\ and\ not\ detected)}{\widehat{P}(not\ detected)} \\ =& \frac{\widehat{\psi}(1-\widehat{p})^5}{1-\widehat{\psi}[1-(1-\widehat{p})^5]} \\ =& \frac{0.59(1-.26)^5}{1-0.59[1-(1-0.26)^5]} = 0.25\end{aligned}$$

Single Species; Single-Season - PRESENCE - A new model

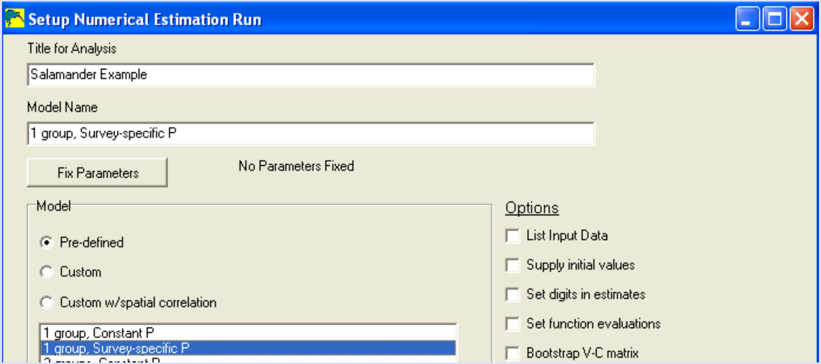
We assumed that p was equal across all sampling occasions. But suppose that effort/ conditions varied across occasions. Need to relax the assumption that p was equal across all occasions.

Standard Notation:

- Model $\psi(*), p(*)$ had equal p across all ocasions.
- Model $\psi(*), p(t)$ has DIFFERENT p for each occasion.

Single Species; Single-Season - $\psi(*), p(t)$

Go back to PRESENCE, fit a new model:



Setup Numerical Estimation Run

Title for Analysis: Salamander Example

Model Name: 1 group, Survey-specific P

No Parameters Fixed

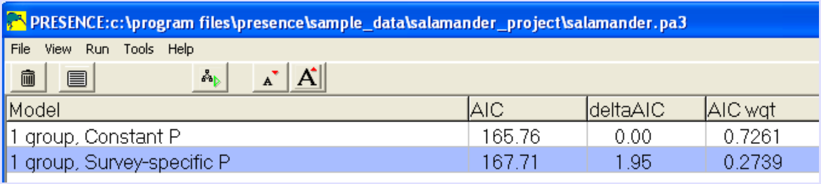
Model
- ● Pre-defined
- ○ Custom
- ○ Custom w/spatial correlation

1 group, Constant P
1 group, Survey-specific P

Options
- ☐ List Input Data
- ☐ Supply initial values
- ☐ Set digits in estimates
- ☐ Set function evaluations
- ☐ Bootstrap V-C matrix

Single Species; Single-Season - $\psi(*)$, $p(t)$

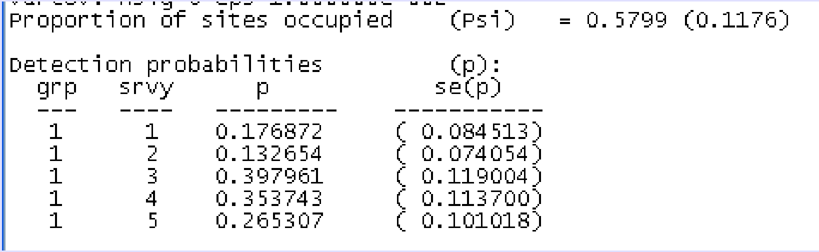
It is added to the RESULTS window:



Model	AIC	deltaAIC	AIC wgt
1 group, Constant P	165.76	0.00	0.7261
1 group, Survey-specific P	167.71	1.95	0.2739

Single Species; Single-Season - $\psi(*), p(t)$

Look at the estimates:

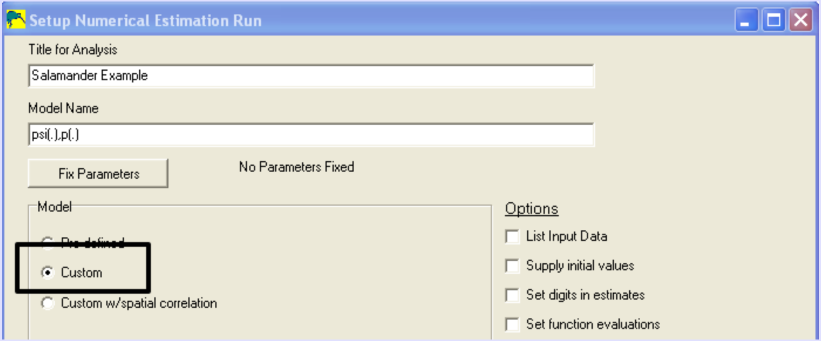


```
Proportion of sites occupied    (Psi)   = 0.5799 (0.1176)

Detection probabilities          (p):
  grp    srvy       p               se(p)
  ---    ----   ----------      ------------
   1       1    0.176872        ( 0.084513)
   1       2    0.132654        ( 0.074054)
   1       3    0.397961        ( 0.119004)
   1       4    0.353743        ( 0.113700)
   1       5    0.265307        ( 0.101018)
```

$\widehat{\psi} = 0.58(SE\ 0.12)$ and each occasion has a separate $\widehat{p}_i$.

Single Species; Single-Season - Custom Models

Refer back to estimates of p_i. Perhaps effort was same in first two occasions and last three occasions. How is a more flexible model fit?



Title for Analysis: Salamander Example

Model Name: psi(.),p(.)

No Parameters Fixed

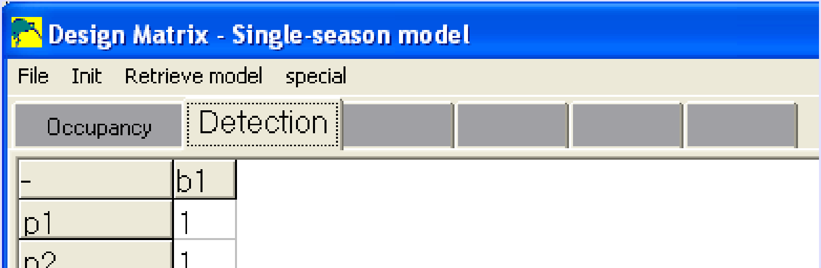
Model
- ○ Pre-defined
- ● Custom
- ○ Custom w/spatial correlation

Options
- ☐ List Input Data
- ☐ Supply initial values
- ☐ Set digits in estimates
- ☐ Set function evaluations

Single Species; Single-Season - Custom Models

This brings up DESIGN MATRIX window, one for each parameter

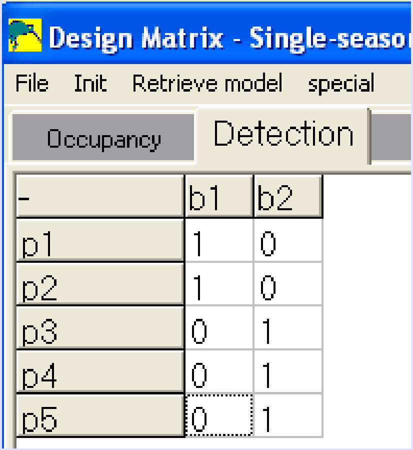
-	a1
psi	1



-	b1
p1	1
p2	1

Single Species; Single-Season - Custom Models

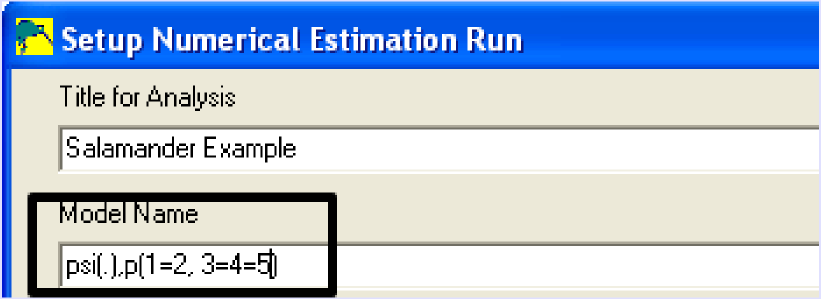
“Real” parameters (rows) vs “Beta” parameters (cols). You want a distinct column for each separate parameter for the detection rates. [Try right clicking.] Get the following structure:



-	b1	b2
p1	1	0
p2	1	0
p3	0	1
p4	0	1
p5	0	1

Single Species; Single-Season - Custom Models

Then run the model with an appropriate name for the model:



Title for Analysis

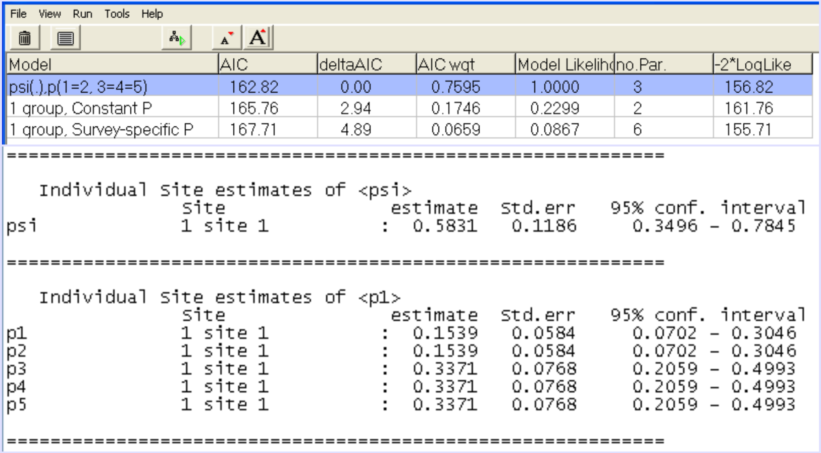
Salamander Example

Model Name

psi(.),p(1=2, 3=4=5)

Single Species; Single-Season - Custom Models

Model is added to results table, and look at estimates



Model	AIC	deltaAIC	AIC wgt	Model Likelih	no.Par.	-2*LogLike
psi(.),p(1=2, 3=4=5)	162.82	0.00	0.7595	1.0000	3	156.82
1 group, Constant P	165.76	2.94	0.1746	0.2299	2	161.76
1 group, Survey-specific P	167.71	4.89	0.0659	0.0867	6	155.71

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

   Individual Site estimates of <psi>
                Site                  estimate   Std.err    95% conf. interval
psi             1 site 1            :  0.5831    0.1186      0.3496 - 0.7845

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

   Individual Site estimates of <p1>
                Site                  estimate   Std.err    95% conf. interval
p1              1 site 1            :  0.1539    0.0584      0.0702 - 0.3046
p2              1 site 1            :  0.1539    0.0584      0.0702 - 0.3046
p3              1 site 1            :  0.3371    0.0768      0.2059 - 0.4993
p4              1 site 1            :  0.3371    0.0768      0.2059 - 0.4993
p5              1 site 1            :  0.3371    0.0768      0.2059 - 0.4993

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

Single Species; Single-Season - Custom Models

Create and fit. a model where $p_1 = p_2$, $p_3 = p_4$, and p_5 is separate.

Single Species; Single-Season - AIC

Which model is best?

“All models are wrong, but some are useful” (George Box)

AIC = Akaike Information Criteria = tradeoff between fit and complexity

$$AIC = -2 \log likelihood + 2\#p$$

where $\#p$ is the number of parameters.

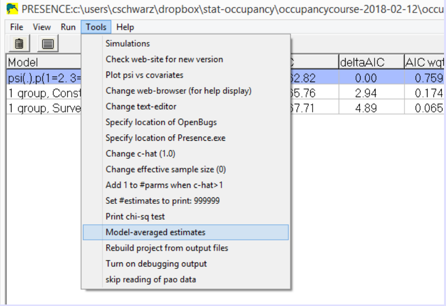
“Best” model in the set is one with arithmetically smallest AIC. Find

$$\Delta AIC_m = AIC_m - AIC_{min}$$

Differences of 2 or 3 AIC units imply no real preference.

Single Species; Single-Season - AIC

Back to RESULTS window. Which model is “best” within model set?



PRESENCE:c:\users\cschwarz\dropbox\stat-occupancy\occupancycourse-2018-02-12\occu

Model	C	deltaAIC	AIC wgt
psi(.),p(1=2. 3=	2.82	0.00	0.759
1 group, Const	5.76	2.94	0.174
1 group, Surve	7.71	4.89	0.065

This creates a text file in your directory with results.

Unfortunately, this is still very much in *beta testing* mode and does not always work as advertised.

Single Species; Single-Season - *RPresence*

RPresence is an *R* package that is able to fit some of the simpler models available in *PRESENCE*.

Key advantages are:

- Scripts so that analyzes can be reused.
- No more clicking
- Much easier extracting output
- Available for Macintosh, Windoze, and Linux platforms.

Key disadvantages are:

- Not all analyses can be done in *RPresence*.

Single Species; Single-Season - *RPresence*

Installation of package (not available on CRAN)

- Download package file from Patuxent software package
- Install from *Package Archive File*

Open the sample program in the *Salamander* directory

Single Species; Single-Season - *RPresence* - Salamander

Basic steps in analysis:

- Input the history file and create the **.pao* file
- Fit some model using the *occMod()* function.
- Model average using the *createAicTable* function.
- Extract results and plot results etc

Refer to sample analysis script.

Single Species; Single-Season - *RPresence* - Salamander

Read in raw data and so some basic error checking:

```
input.data <- readxl::read_excel(file.path("..","salamande
                                 sheet="CompleteData",
                                 col_names=FALSE)   # notic

# do some basic checks on your data
nrow(input.data)
ncol(input.data)
head(input.data)
```

Histories is a matrix ($n_{sites} \times n_{visits}$) of 1's, 0's, or NA's.

Single Species; Single-Season - *RPresence* - Salamander

Create the **.pao* file.

```
salamander.pao <- createPao(input.data,
                            title='Salamander SSSS')
salamander.pao
```

Single Species; Single-Season - *RPresence* - Salamander

Fitting a model *psi(.), p(.)*

```
mod.pdot <- occMod(model=list(psi~1, p~1),
           type="so",
           data=salamander.pao)
```

Note that formula DO NOT HAVE AN = SIGN.

This creates an object with MANY slots and extractor functions.

Single Species; Single-Season - *RPresence* - Salamander

Lot of interesting stuff!

- Basic statistics about your data - make sure these are correct
- Information about fitting the model - typically not of interest.
- *Beta* estimates - typically not of interest
- Estimates of occupancy and detectability.
- Posterior probability of occupancy.

Single Species; Single-Season - *RPresence* - Salamander

The estimates for ψ are given for EVERY site. In this case, there are no covariates so every site has the same occupancy probability.

```
mod.pdot.psi <-mod.pdot$real$psi[1,]   # occupancy probabil
mod.pdot.psi
```

```
            est        se lower_0.95 upper_0.95
unit1 0.5946225 0.1225937  0.3512137  0.7989789
```

Single Species; Single-Season - *RPresence* - Salamander

The estimates for *p* are given for EVERY site-survey combination. In this case, all are the same over all sites and surveys.

```
mod.pdot.p    <- mod.pdot$real$p[seq(1, by=nrow(input.data)
mod.pdot.p
```

```
> mod.pdot.p
                est        se lower_0.95 upper_0.95
unit1_1-1 0.258729 0.0576996  0.1621604  0.3862912
unit1_1-2 0.258729 0.0576996  0.1621604  0.3862912
unit1_1-3 0.258729 0.0576996  0.1621604  0.3862912
unit1_1-4 0.258729 0.0576996  0.1621604  0.3862912
unit1_1-5 0.258729 0.0576996  0.1621604  0.3862912
```

Single Species; Single-Season - *RPresence* - Salamander

Finally, the posterior probability of occupancy

```
> mod.pdot$derived$psi_c
             est        se lower_0.95 upper_0.95
unit1  1.0000000 0.0000000 1.00000000     1.0000
unit2  1.0000000 0.0000000 1.00000000     1.0000
...
unit15 0.2471562 0.1471721 0.06512707     0.6074
unit16 0.2471562 0.1471721 0.06512707     0.6074
...
```

Single Species; Single-Season - *RPresence* - Salamander

Fit a model where detection varies across visits. Notice the use of the reserved keyword *SURVEY* to represent the visits.

```
mod.pt <- occMod(model=list(psi~1, p~SURVEY),
                 type="so",
                 data=salamander.pao)
summary(mod.pt)

mod.pt$real$psi[1,]
mod.pt$real$p[seq(1, by=nrow(input.data), length.out=ncol(
```

Single Species; Single-Season - *RPresence* - Salamander

Fit a model where $p_1 = p_2; p_3 = p_4 = p_5$.
You need to create a data frame which has $n_{sites} \times n_{visits}$ rows.
The ordering is covariates for visit 1, for all sites; then for visit 2 for all sites; etc.
Create a character covariate for the two detection probabilties.

```
survey.cov <- data.frame(site=rep(1:nrow(input.data),
                              ncol(input.data)),
                         visit=rep(1:ncol(input.data),
                              each=nrow(input.data)),
                         d=rep( c("d1","d1","d2","d2","d2"
                              each=nrow(input.data)))
head(survey.cov)
```

Single Species; Single-Season - *RPresence* - Salamander

This gives the following data.frame. Notice the name of the covariate is *d* which will then be used in the model statement.

```
> survey.cov[c(1:4, 37:41),]
   site visit  d
1     1     1 d1
2     2     1 d1
...

38   38     1 d1
39   39     1 d1
40    1     2 d1
41    2     2 d1
...
```

Single Species; Single-Season - *RPresence* - Salamander

Fit a model where $p_1 = p_2; p_3 = p_4 = p_5$

```
mod.pcustom <- occMod(model=list(psi˜1, p˜d),
                      cov.list=list(p.cov=survey.cov),
                      type="so",
                      data=salamander.pao)

mod.pcustom$real$psi[1,]
mod.pcustom$real$p[seq(1, by=nrow(input.data), length.out=
```

Single Species; Single-Season - *RPresence* - Salamander

Fit a model where $p_1 = p_2; p_3 = p_4 = p_5$

```
> mod.pcustom$real$psi[1,]
            est        se lower_0.95 upper_0.95
unit1 0.5831459 0.1186453  0.3495736  0.7845389

> mod.pcustom$real$p[seq(1, by=nrow(input.data), length.out
                est         se lower_0.95 upper_0.95
unit1_1-1 0.1538956 0.05838078 0.07023254  0.3045734
unit1_1-2 0.1538956 0.05838078 0.07023254  0.3045734
unit1_1-3 0.3371048 0.07678979 0.20591460  0.4993209
unit1_1-4 0.3371048 0.07678979 0.20591460  0.4993209
unit1_1-5 0.3371048 0.07678979 0.20591460  0.4993209
```

Single Species; Single-Season - *RPresence* - Salamander

Model averaging:

```
models<-list(mod.pdot,mod.pt,mod.pcustom)
results<-createAicTable(models)
summary(results)

modAvg(results, param="psi")
```

Single Species; Single-Season - *RPresence* - Salamander

Model averaged results for ψ.

```
> summary(results)
             Model DAIC   wgt npar neg2ll warn.conv warn.VC
1        psi()p(d) 0.00 0.760    3 156.82         0       0
2         psi()p() 2.94 0.175    2 161.76         0       0
3 psi()p(SURVEY) 4.90 0.066    6 155.71         0       0
>
> modAvg(results, param="psi")[1,]
            est        se lower_0.95 upper_0.95
unit1 0.5849351 0.1193595  0.3496677   0.786949
```

Single Species; Single-Season - *RPresence* - Salamander

Model averaging - a more general approach
Open the *salamander-modelavg-general.R* file.

- Read in the data
- Create the history data frame
- Create the survey covariate data frame
- Create the *.pao object

Single Species; Single-Season - *RPresence* - Salamander

Model averaging - a more general approach
Create a list of model to be fit.

```
# Notice the commas between the column and the placement o
model.list.csv <- textConnection("
p,               psi
~1,              ~1
~SURVEY,        ~1
~d,             ~1")

model.list <- read.csv(model.list.csv, header=TRUE, as.is=
model.list
```

Single Species; Single-Season - *RPresence* - Salamander

Model averaging - a more general approach

Fit the set of model

```
model.fits <- plyr::alply(model.list, 1, function(x,detect
  cat("\n\n***** Starting ", unlist(x), "\n")
  fit <- RPresence::occMod(model=list(as.formula(paste("ps
                                      as.formula(paste("p"
                             data=detect.pao,type="so")
  fit
},detect.pao=salamander.pao)
```

Single Species; Single-Season - *RPresence* - Salamander

Model averaging - a more general approach

- Notice how you can inspect the output from a particular model
- Model average in the usual way.

Single Species; Single-Season - *RPresence* - Salamander

Summary:

- Don't put = signs in formula
- Creating visit covariates is a bit tricky
- Same output as in *PRESENCE* and *MARK*.
- Also able to read and write *PRESENCE* files directly.

Single Species; Single-Season - MARK vs PRESENCE

MARK is an alternative program to PRESENCE.

Choice between MARK and PRESENCE:

- About 70% of models are common across both platforms.
- Each program has some specialized models (e.g. PRESENCE has spatial and mixture models)
- MARK handles model selection "better"

Cautions about switching between platforms:

- Similar (but not the same) data structures.
- Model specified differently.

Single Species; Single-Season - MARK - Input file format

MARK input file format:

- *.inp file name but a simple text file
- each line take format

  ```
  hhhhh count ; /* This is a comment */
  ```

  where h is 0 or 1; count is typically 1; add comments anywhere.
- Example

  ```
  10100 1; /* site 27 */
  11100 1; /* site 33 */
  00000 1; /* site 45 */
  ```

Single Species; Single-Season - MARK - Salamander

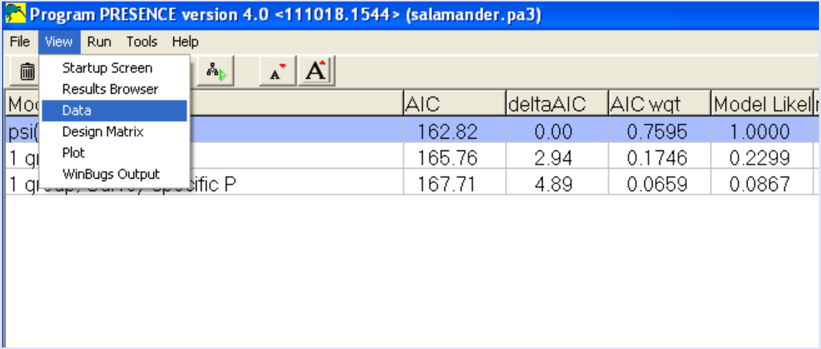
Converting from PRESENCE data.
Open PRESENCE and find old project or open new project

Ver 4.0 (6Oct2011) - Added models: multi-season-het, multi-season-2species

Ver 3.1 (14Oct2010) - Added model: single-season-false-positive-detections

Single Species; Single-Season - MARK - Salamander

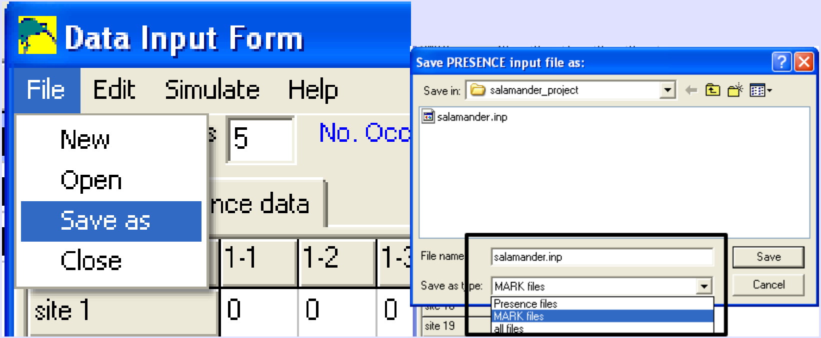
View the data



Mod	AIC	deltaAIC	AIC wgt	Model Likel
psi(	162.82	0.00	0.7595	1.0000
1 gr	165.76	2.94	0.1746	0.2299
1 gr ific P	167.71	4.89	0.0659	0.0867

Single Species; Single-Season - MARK - Salamander

Save the data as a MARK *.inp file

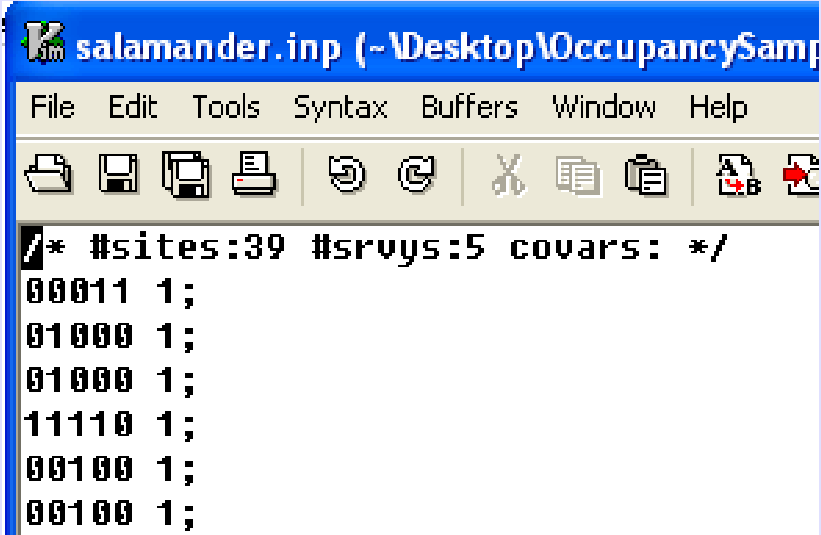


File name: salamander.inp

Save as type: MARK files

Single Species; Single-Season - MARK - Salamander

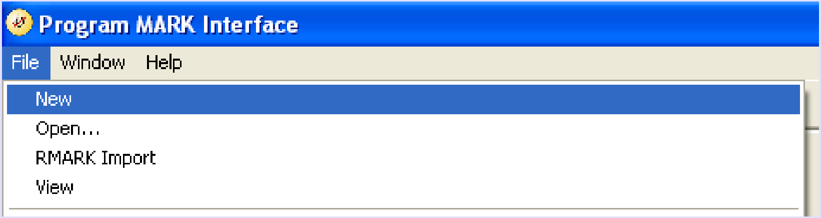
View the MARK *.inp file



```
/* #sites:39 #srvys:5 covars: */
00011 1;
01000 1;
01000 1;
11110 1;
00100 1;
00100 1;
```

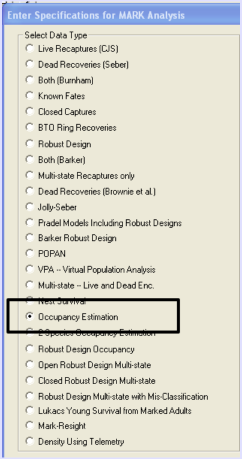
Single Species; Single-Season - MARK - Salamander

Launch MARK and select NEW project.



Single Species; Single-Season - MARK - Salamander

Select class of models (Occupancy models)



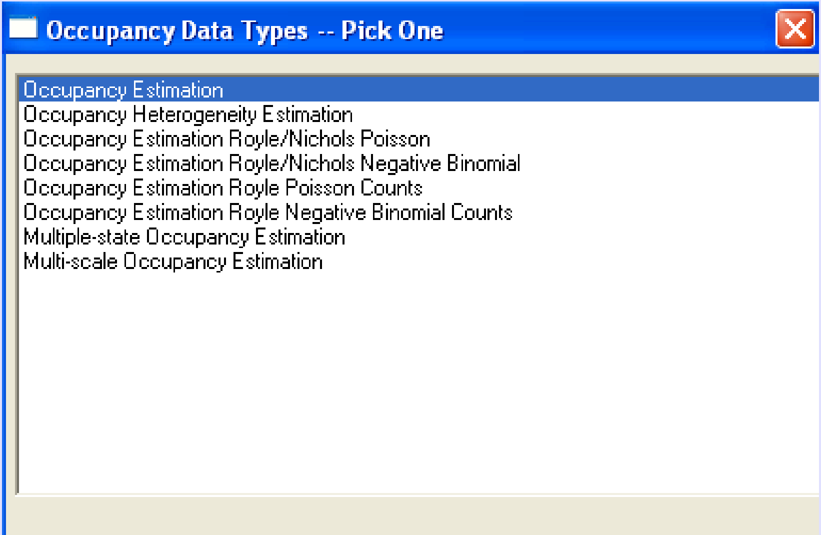
Enter Specifications for MARK Analysis

Select Data Type

- ○ Live Recaptures (CJS)
- ○ Dead Recoveries (Seber)
- ○ Both (Burnham)
- ○ Known Fates
- ○ Closed Captures
- ○ BTO Ring Recoveries
- ○ Robust Design
- ○ Both (Barker)
- ○ Multi-state Recaptures only
- ○ Dead Recoveries (Brownie et al.)
- ○ Jolly-Seber
- ○ Pradel Models Including Robust Designs
- ○ Barker Robust Design
- ○ POPAN
- ○ VPA -- Virtual Population Analysis
- ○ Multi-state -- Live and Dead Enc.
- ○ Nest Survival
- ● Occupancy Estimation
- ○ 2 Species Occupancy Estimation
- ○ Robust Design Occupancy
- ○ Open Robust Design Multi-state
- ○ Closed Robust Design Multi-state
- ○ Robust Design Multi-state with Mis-Classification
- ○ Lukacs Young Survival from Marked Adults
- ○ Mark-Resight
- ○ Density Using Telemetry

Single Species; Single-Season - MARK - Salamander

Select type of Occupancy Model

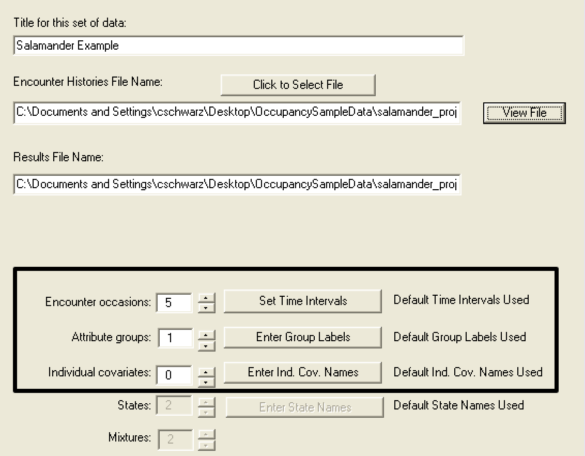


Occupancy Data Types -- Pick One

Occupancy Estimation
Occupancy Heterogeneity Estimation
Occupancy Estimation Royle/Nichols Poisson
Occupancy Estimation Royle/Nichols Negative Binomial
Occupancy Estimation Royle Poisson Counts
Occupancy Estimation Royle Negative Binomial Counts
Multiple-state Occupancy Estimation
Multi-scale Occupancy Estimation

Single Species; Single-Season - MARK - Salamander

Complete the initial screen. CAUTION - you cannot correct errors after you finish this screen. MARK creates specialized files (*.FPT, *.DBF) - DO NOT CHANGE.



Title for this set of data: Salamander Example

Encounter Histories File Name: C:\Documents and Settings\cschwarz\Desktop\OccupancySampleData\salamander_proj

Results File Name: C:\Documents and Settings\cschwarz\Desktop\OccupancySampleData\salamander_proj

Encounter occasions: 5 — Default Time Intervals Used

Attribute groups: 1 — Default Group Labels Used

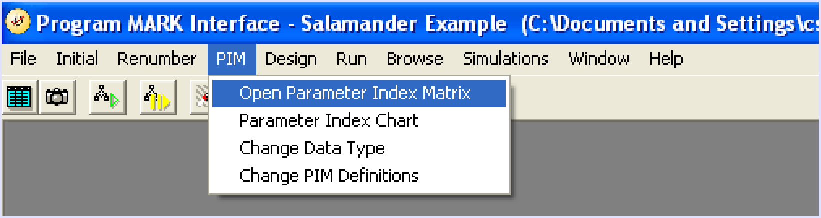
Individual covariates: 0 — Default Ind. Cov. Names Used

States: 2 — Default State Names Used

Mixtures: 2

Single Species; Single-Season - MARK - Salamander

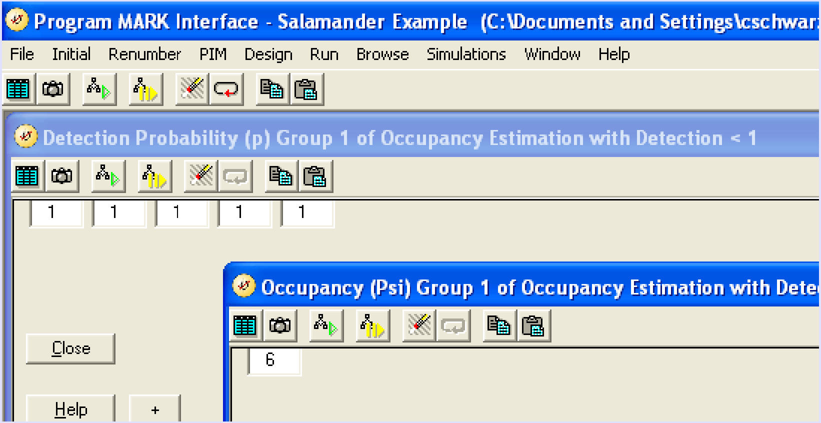
Now you are ready to specify model - hurray!
MARK used a different (but similar) paradigm to PRESENCE.
Open the PIM menu:



Single Species; Single-Season - MARK - Salamander

Specify entries in PIM to make parameters equal. For example the $\psi(*), p(*)$ model has entries below. The actual values are NOT important.

CAUTION: Do NOT share PIM entries across parameters.

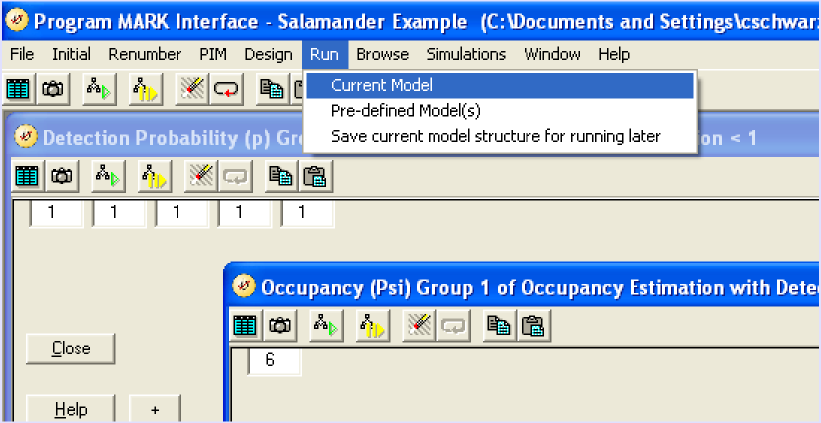


1	1	1	1	1

6

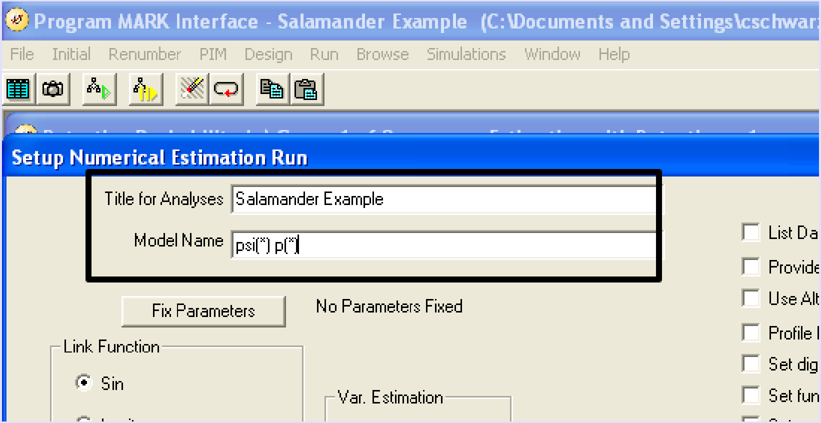
Single Species; Single-Season - MARK - Salamander

RUN the model



Single Species; Single-Season - MARK - Salamander

Specify a sensible model name. Use the standard naming conventions.



Title for Analyses: Salamander Example

Model Name: psi(*) p(*)

No Parameters Fixed

● Sin

☐ List Da

☐ Provide

☐ Use Alt

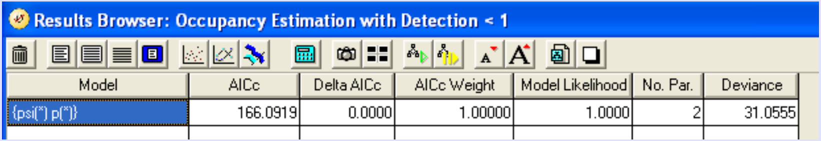
☐ Profile I

☐ Set dig

☐ Set fun

Single Species; Single-Season - MARK - Salamander

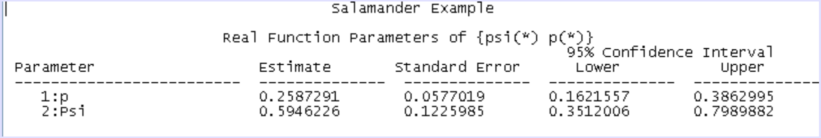
Look at RESULTS Window. Similar to PRESENCE but AIC has a second order correction factor applied so slightly different from PRESENCE AIC.



Model	AICc	Delta AICc	AICc Weight	Model Likelihood	No. Par.	Deviance
{psi(*) p(*)}	166.0919	0.0000	1.00000	1.0000	2	31.0555

Single Species; Single-Season - MARK - Salamander

Look at REAL estimates. Same as PRESENCE (not surprising)



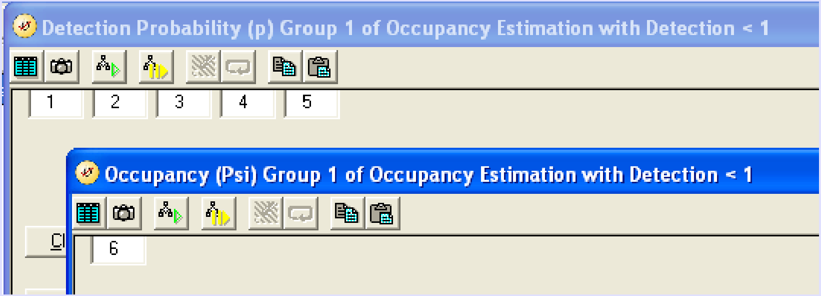
```
                                   Salamander Example

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

Single Species; Single-Season - MARK - Salamander

Fit the model where p varies over sampling occasions.
Fit the model where $p(1) = p(2)$ and $p(3) = p(4) = p(5)$.

Single Species; Single-Season - MARK - Salamander

PIM for model where *p* varies over sampling occasions.



Detection Probability (p) Group 1 of Occupancy Estimation with Detection < 1

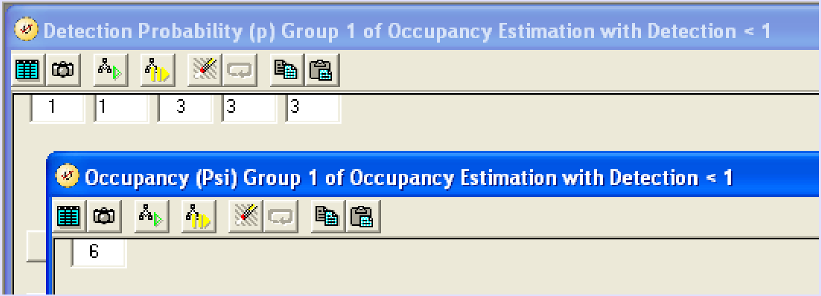
1	2	3	4	5

Occupancy (Psi) Group 1 of Occupancy Estimation with Detection < 1

6

Single Species; Single-Season - MARK - Salamander

PIM for model where $p(1) = p(2)$ and $p(3) = p(4) = p(5)$.



Detection Probability (p) Group 1 of Occupancy Estimation with Detection < 1

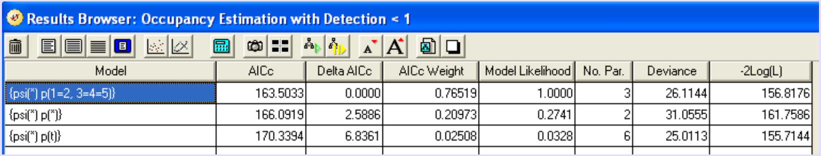
1	1	3	3	3

Occupancy (Psi) Group 1 of Occupancy Estimation with Detection < 1

6

Single Species; Single-Season - MARK - Salamander

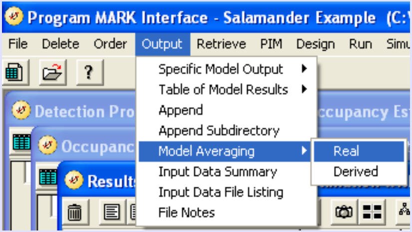
RESULTS window for 3 models. Think of relative support for these 3 models and the implications of the results.

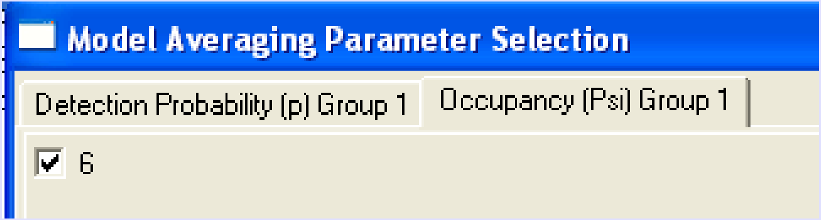


Model	AICc	Delta AICc	AICc Weight	Model Likelihood	No. Par.	Deviance	-2Log(L)
{psi(*) p(1=2, 3=4=5)}	163.5033	0.0000	0.76519	1.0000	3	26.1144	156.8176
{psi(*) p(*)}	166.0919	2.5886	0.20973	0.2741	2	31.0555	161.7586
{psi(*) p(t)}	170.3394	6.8361	0.02508	0.0328	6	25.0113	155.7144

Single Species; Single-Season - MARK - Salamander

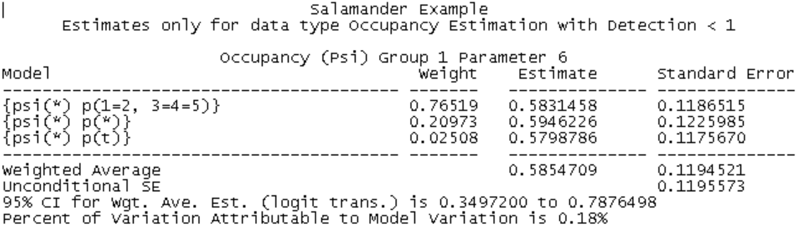
Model average the $\widehat{\psi}$.





Single Species; Single-Season - MARK - Salamander

Results of model averaging for ψ.



```
|                              Salamander Example
     Estimates only for data type Occupancy Estimation with Detection < 1

                        Occupancy (Psi) Group 1 Parameter 6
Model                                    Weight     Estimate          Standard Error
--------------------------------------   -------    --------------    --------------
{psi(*) p(1=2, 3=4=5)}                   0.76519    0.5831458         0.1186515
{psi(*) p(*)}                            0.20973    0.5946226         0.1225985
{psi(*) p(t)}                            0.02508    0.5798786         0.1175670
--------------------------------------   -------    --------------    --------------
Weighted Average                                    0.5854709         0.1194521
Unconditional SE                                                      0.1195573
95% CI for Wgt. Ave. Est. (logit trans.) is 0.3497200 to 0.7876498
Percent of Variation Attributable to Model Variation is 0.18%
```

Because estimates are so similar across models, the model averaged estimate is not that different and the model averaged se (unconditional SE in output) is only slightly larger than individual se.

Single Species; Single-Season - Presence

Single-Species Single-Season Occupancy Studies

Using *RMark*

Single Species; Single-Season - *RMark*

RMark is an *R* package that interfaces with *MARK*
Key advantages are:

- Scripts so that analyzes can be reused.
- No more clicking
- Much easier extracting output

Key disadvantages are:

- Only available on Windoze platforms.

Single Species; Single-Season - *RMark*

Installation of package

- Download and Install package file from CRAN in the usual way

Open the sample program in the *Salamander* directory

Single Species; Single-Season - *RMark* - Salamander

Basic steps in analysis:

- Input the occupancy history and *process* the data.
- Fit some model using the *mark* function.
- Extract information from returned object
- Model average using the *collect.models* function.
- Extract results and plot results etc

Refer to sample analysis scripts.

Single Species; Single-Season - *RMark* - Salamander

Read in raw data and so some basic error checking:

```
input.data <- readxl::read_excel(file.path("..","salamande
                                 sheet="CompleteData",
                                 col_names=FALSE)   # notic
```

Histories is a matrix ($n_{sites} \times n_{visits}$) of 1's, 0's, or NA's.

Single Species; Single-Season - *RMark* - Salamander

Basic data structure is a data.frame with the following elements

- *ch* - capture-history - character vector where each element is a capture-history in *MARK* format, i.e. "0111".
  Missing visits indicated by a period (.)
- *freq* - frequency of occurrence (usually all 1's)
- *other variables* - site-level covariates

Single Species; Single-Season - *RMark* - Salamander

Basic data structure is a data.frame with the following elements

```
# Extract the history records and create a capture history
input.history <- data.frame(freq=1,
                            ch=apply(input.data[,1:5],1,pa
head(input.history)

# Change any NA to . in the chapter history
input.history$ch <- gsub("NA",".", input.history$ch, fixed
```

Single Species; Single-Season - *RMark* - Salamander

Process the data

```
sal.data <- process.data(data=input.history,
                          model="Occupancy")

summary(sal.data)
```

Single Species; Single-Season - *RMark* - Salamander

Fitting a model *psi(.), p(.)*

```
# What are the parameter names for Single Season Single Sp
setup.parameters("Occupancy", check=TRUE)

mod.fit <-  RMark::mark(sal.data,
                        model="Occupancy",
                        model.parameters=list(
                          Psi   =list(formula=~1),
                          p     =list(formula=~1)
                        )
                     )
summary(mod.fit)
```

This creates an object with MANY sub-objects..

Single Species; Single-Season - *RMark* - Salamander

Lot of interesting stuff!

- Basic statistics about your data - make sure these are correct
- Information about fitting the model - typically not of interest.
- *Beta* estimates - typically not of interest
- *real* estimates - usually of primary interest
- Estimates of occupancy and detectability.
- Posterior probability of occupancy.

Single Species; Single-Season - *RMark* - Salamander

The estimates for ψ can be found in several ways:

```
mod.fit$results$real# on the regular 0-1 scale for each si

get.real(mod.fit, "Psi", se=TRUE)
```

```
             all.diff.index par.index  estimate        se
Psi g1 a0 t1              6         2 0.5946226 0.1225985 

                ucl fixed    note group age time Age Ti
Psi g1 a0 t1 0.7989882                   1   0    1   0
```

Single Species; Single-Season - *RMark* - Salamander

The estimates for *p* found in a similar fashion

```
get.real(mod.fit, "p", se=TRUE)
```

```
           all.diff.index par.index  estimate        se
p g1 a0 t1              1         1 0.2587291 0.0577019 0.
p g1 a1 t2              2         1 0.2587291 0.0577019 0.
p g1 a2 t3              3         1 0.2587291 0.0577019 0.
p g1 a3 t4              4         1 0.2587291 0.0577019 0.
p g1 a4 t5              5         1 0.2587291 0.0577019 0.

                 ucl fixed    note group age time Age Tim
p g1 a0 t1 0.3862995                     1   0    1   0    0
p g1 a1 t2 0.3862995                     1   1    2   1    1
p g1 a2 t3 0.3862995                     1   2    3   2    2
p g1 a3 t4 0.3862995                     1   3    4   3    3
p g1 a4 t5 0.3862995                     1   4    5   4    4
```

Single Species; Single-Season - *RMark* - Salamander

The posterior probability of occupancy is NOT available but can be computed by hand if needed - contact me.

Single Species; Single-Season - *RMark* - Salamander

Fit a model where detection varies across visits.
Replace $p \sim 1$ by $p \sim time$.
Note *time* and *Time* are reserved keywords.

- *time* implies a separate capture probability for each visit
- *Time* implies a linear change in capture-probability over visits

Fit the new model and look at resutls

Single Species; Single-Season - *RMark* - Salamander

Fit a model where $p_1 = p_2; p_3 = p_4 = p_5$.
This is a SURVEY-LEVEL covariate and so you need to modify the *ddl* and pass the revised *ddl* to *RMark*. [Refer to discussion of covariates later in the course.]

```
sal.ddl <- make.design.data(sal.data)
sal.ddl

sal.ddl$p$Effort <- factor(c("d1","d1","d2","d2","d2"))
sal.ddl
```

Single Species; Single-Season - *RMark* - Salamander

Fit a model where $p_1 = p_2; p_3 = p_4 = p_5$.
This gives the following *ddl* for p

```
  par.index model.index group age time Age Time Effort
1         1           1     1   0    1   0    0     d1
2         2           2     1   1    2   1    1     d1
3         3           3     1   2    3   2    2     d2
4         4           4     1   3    4   3    3     d2
5         5           5     1   4    5   4    4     d2
```

Single Species; Single-Season - *RMark* - Salamander

Fit a model where $p_1 = p_2; p_3 = p_4 = p_5$
Add the revised *ddl* to the fit

```
mod.fit3 <-  RMark::mark(sal.data, ddl=sal.ddl,
                         model="Occupancy",
                         model.parameters=list(
                           Psi   =list(formula=~1),
                           p     =list(formula=~Effort)
                         )
)
```

Single Species; Single-Season - *RMark* - Salamander

Fit a model where $p_1 = p_2; p_3 = p_4 = p_5$

```
              all.diff.index par.index  estimate        se
p g1 a0 t1                 1         1 0.1538956 0.0583875 0.0
p g1 a1 t2                 2         1 0.1538956 0.0583875 0.0
p g1 a2 t3                 3         2 0.3371047 0.0767915 0.2
p g1 a3 t4                 4         2 0.3371047 0.0767915 0.2
p g1 a4 t5                 5         2 0.3371047 0.0767915 0.2
                 ucl fixed    note group age time Age Time
p g1 a0 t1 0.3045984                     1   0    1   0    0
p g1 a1 t2 0.3045984                     1   1    2   1    1
p g1 a2 t3 0.4993277                     1   2    3   2    2
p g1 a3 t4 0.4993277                     1   3    4   3    3
p g1 a4 t5 0.4993277                     1   4    5   4    4
```

Single Species; Single-Season - *RMark* - Salamander

Model averaging:

```
model.set <- RMark::collect.models( type="Occupancy")
model.set
```

```
                   model npar     AICc DeltaAICc    weight Devia
3 p(~Effort)Psi(~1)    3 163.5033  0.000000 0.7651935 26.1
1      p(~1)Psi(~1)    2 166.0919  2.588649 0.2097265 31.0
2   p(~time)Psi(~1)    6 170.3394  6.836116 0.0250800 25.0
```

Single Species; Single-Season - *RMark* - Salamander

Model averaged results for ψ.

```
Psi.ma <- RMark::model.average(model.set, param="Psi")
Psi.ma
```

```
             par.index  estimate        se fixed    note g
Psi g1 a0 t1         6 0.5854709 0.1195573
             Age Time
Psi g1 a0 t1   0    0
```

Single Species; Single-Season - *RMark* - Salamander

Model averaging - a more general approach.

Open the *salamander-modelavg-general.R* file.

- Read in the data
- Create the history data frame
- Modify the *ddl* for when need survey level covariates
- Process the data
- Read in the models to be fit
- Fit the model set
- Model average

Single Species; Single-Season - *RMark* - Salamander

Model averaging - a more general approach

Create a list of model to be fit. CAUTION - check the parameter names first

```
setup.parameters("Occupancy", check=TRUE)

# Get the list of models
model.list.csv <- textConnection("
p,            Psi
 ~1,         ~1
 ~time,      ~1
 ~Effort,    ~1")

model.list <- read.csv(model.list.csv, header=TRUE, as.is=
model.list$model.number <- 1:nrow(model.list)
model.listmodel.list <- read.csv(model.list.csv, header=TRU
model.list
```

Single Species; Single-Season - *RMark* - Salamander

Model averaging - a more general approach

Fit the set of model

```
model.fits <- plyr::dlply(model.list, "model.number", funct
  cat("\n\n***** Starting ", unlist(x), "\n")

    fit <- RMark::mark(input.data, ddl=input.ddl,
                     model="Occupancy",
                     model.parameters=list(
                       Psi   =list(formula=as.formula(eval
                       p     =list(formula=as.formula(eval
                     )
                     #,brief=TRUE,output=FALSE, delete=TRU
                     #,invisible=TRUE,output=TRUE  # set f
  )
  mnumber <- paste("m...",formatC(x$model.number, width = 
  assign( mnumber, fit, envir=.GlobalEnv)
  #browser()
```

Single Species; Single-Season - *RMark* - Salamander

Model averaging - a more general approach

- Notice how you can inspect the output from a particular model
- Collect models and Model average in the usual way.

Single Species; Single-Season - *RMark* - Salamander

Summary:

- Creating visit covariates is a bit tricky and is done using the *ddl*
- Same result as other packages
- Able to rexport to *MARK* the final set of file (useful for GOF and other features)

Single Species; Single-Season - Grossbeaks

An occupancy study was made on Blue Grosbeaks (*Guiraca caerulea*) on 41 old fields planted to longleaf pines (*Pinus palustris*) in southern Georgia, USA.

Surveys were 500 m transects across each field and were completed three times during the breeding season in 2001.

Columns in the file are:

- *field* - field number
- *v1, v2, v3* - detection histories for each site on each of 3 visit during the 2001 breeding season.
- *field.size* - size of the files
- *bqi* - Enrollment in bobwhite quail initiative; does occupancy increase if field belongs to this initiative?
- *crop.hist* - crop history
- *crop1, crop2* - indicator variables for the crop history
- *count1, count2, count3* - actual counts of birds detected in each visit

Single Species; Single-Season - Exercise 01

Using the package of your choice, fit the following models:

- $\psi(.), p(.)$
- $\psi(.), p(t)$

Obtain model averaged estimates for the parameter ψ.
Save your scripts as we will add more models in future exercises.

Single Species; Single-Season - Missing data

Single-Species Single-Season Occupancy Studies

Missing Data

Single Species; Single-Season - Missing Values

Types of missing values:

- Missing Completely at Random (MCAR)
    - Missingness unrelated to response or covariates
    - No biases introduced; loss of precision; no adjustment needed
- Missing at Random (MAR)
    - Missingness unrelated to response, but may be related to covariates
    - (Mostly) no biases introduced loss or precision; some reweighting may be required
- Informative Missing (IM)
    - Missingness related to response
    - DIFFICULT STATISTICAL PROBLEM!!!

Single Species; Single-Season - Missing Values

Missing Completely at Random:

- Bad weather
- Equipment breakdown

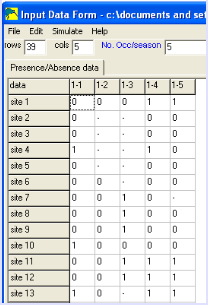
Modify histories in input files as follows:

- *PRESENCE*: Use a "-" (hyphen) to represent a missing value.
- *RPresence*: Use NA to represent a missing value
- *MARK*: Use a "." (period) to represent a missing value.
- *RMark*: Use a "." (period) to represent a missing value in the *ch* variable in the data.frame.
- *unmarked*: Use NA to represent a missing value
- *JAGS*: Need to delete the missing values

Do NOT use a 0 to represent a missing value!

Single Species; Single-Season - Missing Values - PRESENCE

MCAR in Presence:



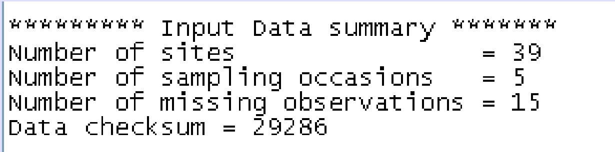
rows 39 cols 5 No. Occ/season 5

data	1-1	1-2	1-3	1-4	1-5
site 1	0	0	0	1	1
site 2	0	-	-	0	0
site 3	0	-	-	0	0
site 4	1	-	-	1	0
site 5	0	-	-	0	0
site 6	0	0	-	0	0
site 7	0	0	1	0	-
site 8	0	0	1	0	0
site 9	0	0	1	0	0
site 10	1	0	0	0	0
site 11	0	0	1	1	1
site 12	0	0	1	1	1
site 13	1	0	-	1	1

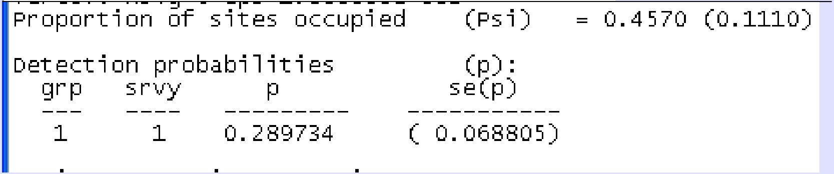
File → Save As *.pao file.
Clear RESULTS window (you can't compare AIC from different data sets).
Run the models as needed.

Single Species; Single-Season - Missing Values - PRESENCE

View estimates file in usual fashion:



```
********** Input Data summary ********
Number of sites                    = 39
Number of sampling occasions       = 5
Number of missing observations = 15
Data checksum = 29286
```



```
Proportion of sites occupied    (Psi)   = 0.4570 (0.1110)

Detection probabilities          (p):
  grp   srvy      p              se(p)
  ---   ----   ----------     ------------
   1      1    0.289734       ( 0.068805)
```

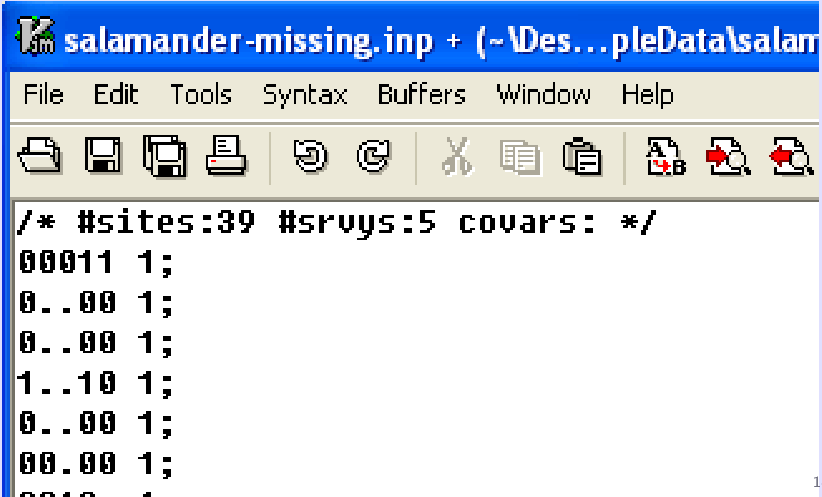
Single Species; Single-Season - Missing Values - *RPresence*

Run *RPresence* in the usual way – refer to *R* code.
Things to watch out for: Reading in '.' or '-' from file for missing values.

```
input.data <- readxl::read_excel("Blue_Ridge_pg99.xls",
                                 sheet="SomeMissingValues"
                                 na='-',
                                 col_names=FALSE)  # notic
```

Single Species; Single-Season - Missing Values - *MARK*

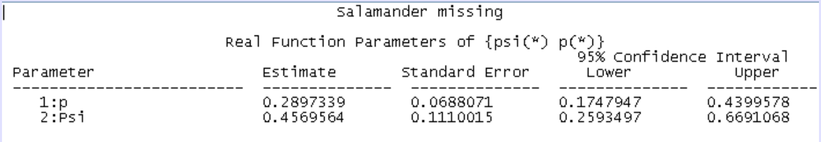
Save as *.inp file and use text editor to change “-” to “.” (Why isn’t this automatic?)



```
/* #sites:39 #srvys:5 covars: */
00011 1;
0..00 1;
0..00 1;
1..10 1;
0..00 1;
00.00 1;
```

Single Species; Single-Season - Missing Values - *MARK*

Run MARK in the usual way – start a new set of *.dbf and *fpt files.



```
                                    Salamander missing

                        Real Function Parameters of {psi(*) p(*)}
                                                                     95% Confidence Interval
Parameter                         Estimate        Standard Error      Lower           Upper
-------------------------------   --------------  --------------  --------------  --------------
   1:p                            0.2897339       0.0688071       0.1747947       0.4399578
   2:Psi                          0.4569564       0.1110015       0.2593497       0.6691068
```

Single Species; Single-Season - Missing Values - *RMark*

Use a period (.) to indicate a missing visit.

```
input.history <- data.frame(freq=1,
                            ch=apply(input.data[,1:5],1,pas
head(input.history)

# Change any NA to . in the chapter history
input.history$ch <- gsub("NA",".", input.history$ch, fixed=
head(input.history)
```

Single Species; Single-Season - Missing Values - *RMark*

Run *RMark* in the usual way. No other changes needed.
Run the sample code and look at the output.

Single Species; Single-Season - Missing Values - MAR

Missing at Random:

- Key assumption is that p_t is same for ALL sites at occasion t.
- But it may be impossible to visit ALL sampled sites at every occasion.
- Visit “odd numbered” sites on visits 1, 3, 5; visit “even numbered” sites on days 2, 4, 6.

No real statistical issues nor problems unless you only visit many sites only once.

Enter data as missing values.

Single Species; Single-Season - Missing Values - IM

Informative missing:

- You have prior knowledge on sites and tend to visit sites that you think are occupied and not visit sites that are not occupied – bad idea as no inference can be made.
- You visit sites ONLY until a detection is observed ("removal design")
  - Model with $p(t)$ cannot be fit! [Parameters are not-identifiable unless you are willing to assume that $p_i = p_j$ for at least one i, j pair.
  - Use a hybrid design where some sites (e.g. 50%) are still revisited even when occupancy has been established.

This type of missing data are very difficult to deal with!

Single Species; Single-Season - Missing Values - Summary

Simplest is dealing with Missing Completely at Random:

- Bad weather
- Equipment breakdown

Modify histories in input files to account for missing observation.

Do NOT use a 0 to represent a missing value!

Single Species; Single-Season

Single-Species Single-Season Occupancy Studies

Goodness-of-fit

Single Species; Single-Season - Model Assessment

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.

If assumptions are not valid then

- Estimators can be biased (either high or low).
- SE are typically too narrow (false degree of confidence in results)

Single Species; Single-Season - Model Assessment

Impact of closure violation, e.g. occupancy changes during the season.

- Random movement -
    - detection rate & movement confounded;
    - Occupancy estimate is unbiased, but now interpreted as "proportion of sites used" rather than occupancy.
- Non-random movement - permanent immigration or emigration
    - For two occasions, ψ estimates initial or final occupancy.
    - For multiple occasions, no simple patterns

Single Species; Single-Season - Model Assessment

Impact of heterogeneity in:

- Occupancy only (e.g. males and females have different occupancy rates; different habitats have different occupancy rates
    - $\widehat{\psi}$ estmates “average” occupancy
    - SE($\widehat{\psi}$) not too bad
    - Mixture models
- Detection rates (e.g. because of animal size, local covariates)
    - $\widehat{\psi}$ tends to be biased downwards; SE are typically too small
    - Measure covariates (global, e.g. temperature at time of survey) or (local, e.g. site-specific features)
    - If detection depends on site abundance, use “abundance” occupancy models.

Single Species; Single-Season - Model Assessment

Impact of non-independence

- Site too small (vocal cues heard in more than one site)
- Detection after first detect easier because you know where animal is located (e.g. nests)
- $\widehat{\psi}$ is often unbiased, but reported SE are too low.

Single Species; Single-Season - Model Assessment

AIC does NOT measure absolute goodness-of-fit and the “best” model of the set may still be a poor fit to the data.

Basic idea is too look at observed and predicted counts. Suppose that $\widehat{\psi} = 0.82$ and $\widehat{p} = 0.43$ for model $\psi(*), p(*)$ from a survey with $s = 50$ sites and $K = 3$ occasions. We observed 10 sites with history 101.

- $E(101) = s \times \widehat{\psi} \times \widehat{p}(1 - \widehat{p})\widehat{p} =$ $50 \times 0.82 \times 0.43(1 - 0.43)0.43 = 4.32$
- This has a “discrepancy” of $\frac{(O_h - E_h)^2}{E_h} = \frac{(10 - 4.32)^2}{4.32} = 7.47$
- $X^2_{gof} = \sum_{all\ \ histories} \frac{(O_h - E_h)^2}{E_h}$
- X^2_{gof} should follow a chi-square distribution and large values of X^2_{gof} indicate lack-of-fit. [This is similar to the deviance column reported in MARK.]
- But because of small counts, need to do a bootstrap

Single Species; Single-Season - Model Assessment

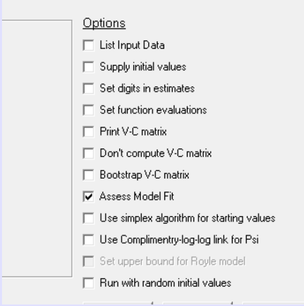
Bootstrapping:

1. Fit model to data
2. Compute measure of lack-of-fit (Deviance) for this model
3. Use the estimates from (2) to “generate” fake data that satisfy the assumptions.
4. Find the Deviance for this fake data.
5. Repeat generating fake data many times (100+)
6. How unusual is the observed deviance from (2) relative to the set of deviances from the fake data in (5)? If very unusual, need to adjust (see later).

Single Species; Single-Season - Model Assessment - *PRESENCE*

A bootstrapping assessment of goodness of fit is available for a limited set of models.

To obtain this check the box for model assessment in the *Run* dialogue:



Options

☐ List Input Data
☐ Supply initial values
☐ Set digits in estimates
☐ Set function evaluations
☐ Print V-C matrix
☐ Don't compute V-C matrix
☐ Bootstrap V-C matrix
☑ Assess Model Fit
☐ Use simplex algorithm for starting values
☐ Use Complimentry-log-log link for Psi
☐ Set upper bound for Royle model
☐ Run with random initial values

Then enter the number of bootstrap samples.

Single Species; Single-Season - Model Assessment - *PRESENCE*

The goodness of fit assessment occurs at the end of the output file - I:

```
Assessing Model Fit for Single-season model:

History(cohort) Observed        Expected        Chi-square
00011( 0 0)          1.0000      0.632306119         0.21
01000( 0 1)          2.0000      1.811586527         0.02
11110( 0 3)          1.0000      0.077030726        11.06
...
00000( 0 14)        21.0000     20.999999940          0.00
00010( 0 35)         2.0000      1.811586527          0.02
00001( 0 37)         2.0000      1.811586527          0.02
Test Statistic =    44.7261 min(expect)=7.703073e-002
```

Single Species; Single-Season - Model Assessment - *PRESENCE*

The goodness of fit assessment occurs at the end of the output file - II:

```
Test Statistic (data)    =         44.7261

From 100 parametric bootstraps...
Probability of test statistic >= observed = 0.0396
Lowest  simulated Test Stat =         10.7058
Average simulated Test Stat =         27.8241
Median simulated Test Stat =         26.5126
Highest simulated Test Stat =         53.0129
Estimate of c-hat = 1.6075   (=TestStat/AvgTestStat)
```

Lack-o-fit is not extreme (p-value of .0396).

Single Species; Single-Season - Model Assessment - *PRESENCE*

The goodness of fit assessment occurs at the end of the output file - III:

```
- Distribution of simulated test statistics ----------
 TestStat  Prop.
  3.9760   0.0000:*
  6.6266   0.0100:*
  9.2773   0.0100:*
 11.9279   0.0100:*
 14.5785   0.0100:*
 17.2292   0.0100:*
 19.8798   0.0400:****
 22.5305   0.1000:*********
 25.1811   0.1600:*************
 27.8318   0.1400:************
```

Single Species; Single-Season - Model Assessment - *PRESENCE*

What causes lack-of-fit?

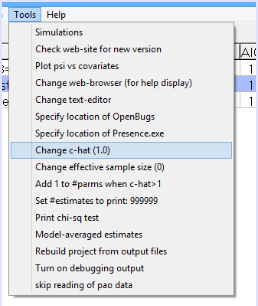
- Structural problems. You fit a $p(*)$ model when p_t varies considerably.
- Heterogeneity. Detection/Occupancy probabilities vary among sites for unknown reasons.
- Non-independence. Detection at one site “influences” detection at another site.

Try to fit more appropriate model (e.g. with covariates).
Minor modification through variance adjustment ($\widehat{c}$). Estimates may still be unbiased, but SE are multiplied by $\sqrt{\widehat{c}}$. [OK if $\widehat{c} < 5$ *or* 6.]
$\widehat{c}$ was estimated above in the model assessment:

```
Estimate of c-hat = 1.6075    (=TestStat/AvgTestStat)
```

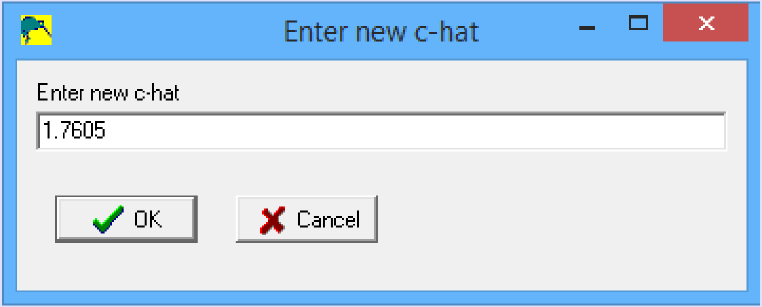
Single Species; Single-Season - Model Assessment - *PRESENCE*

You can adjust the AIC table and estimates for $\widehat{c}$ using:



Single Species; Single-Season - Model Assessment - *PRESENCE*

Enter the new value for $\widehat{c}$



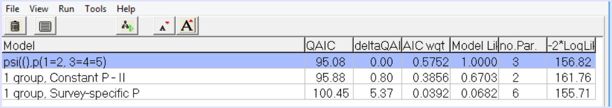
Enter new c-hat

1.7605

Single Species; Single-Season - Model Assessment - *PRESENCE*

This updates the AIC table

The Results window now changes to QAIC etc. Interpreted analogously to AIC.



Model	QAIC	deltaQAI	AIC wgt	Model Li	no.Par.	-2*LogLi
psi(),p(1=2, 3=4=5)	95.08	0.00	0.5752	1.0000	3	156.82
1 group, Constant P - II	95.88	0.80	0.3856	0.6703	2	161.76
1 group, Survey-specific P	100.45	5.37	0.0392	0.0682	6	155.71

Note: *RPresence* does not update the standard errors of the estimates in the output file.

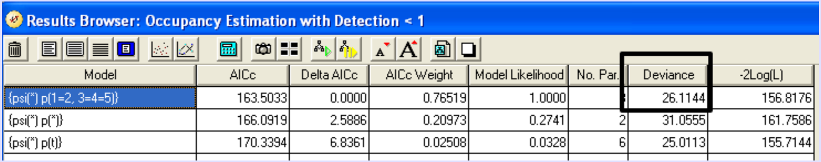
Note: *MARK* DOES update the SEs in the output files.

Single Species; Single-Season - Model Assessment - *RPresence*

Unable to conduct the GOF directly with *RPresence*.
You will need to export the information from *RPresence* to *PRESENCE* and do it in *PRESENCE*.
Contact me for details.

Single Species; Single-Season - Model Assessment - *MARK*

Go back to salamander data and look at results window:



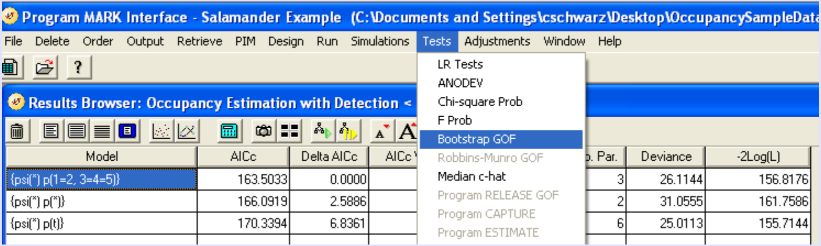
Results Browser: Occupancy Estimation with Detection < 1

Model	AICc	Delta AICc	AICc Weight	Model Likelihood	No. Par.	Deviance	-2Log(L)
{psi(*) p(1=2, 3=4=5)}	163.5033	0.0000	0.76519	1.0000	[illegible]	26.1144	156.8176
{psi(*) p(*)}	166.0919	2.5886	0.20973	0.2741	2	31.0555	161.7586
{psi(*) p(t)}	170.3394	6.8361	0.02508	0.0328	6	25.0113	155.7144

How unusual is a deviance of 26.11?

Single Species; Single-Season - Model Assessment - *MARK*

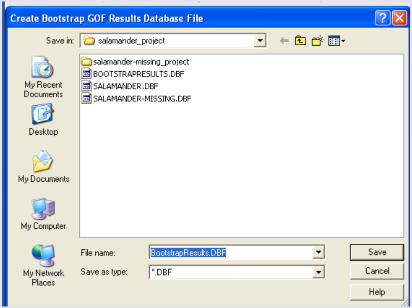
Select Bootstrap GOF from the Tests menu.



Model	AICc	Delta AICc	AICc W	. Par.	Deviance	-2Log(L)
{psi(*) p(1=2, 3=4=5)}	163.5033	0.0000		3	26.1144	156.8176
{psi(*) p(*)}	166.0919	2.5886		2	31.0555	161.7586
{psi(*) p(t)}	170.3394	6.8361		6	25.0113	155.7144

Single Species; Single-Season - Model Assessment - *MARK*

Save the bootstrap results in an appropriate file and generate 100+ fake datasets.



Create Bootstrap GOF Results Database File

Save in: salamander_project

salamander-missing_project
BOOTSTRAPRESULTS.DBF
SALAMANDER.DBF
SALAMANDER-MISSING.DBF

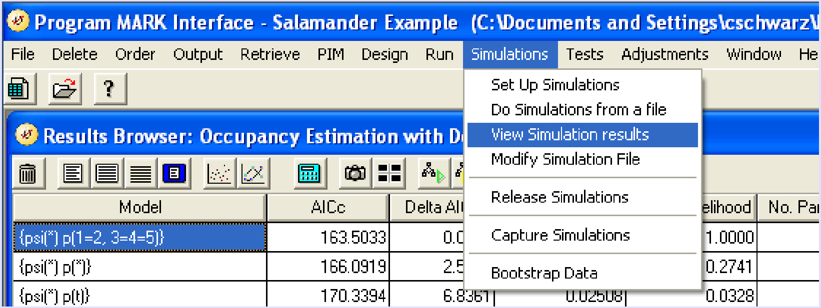
File name: BootstrapResults.DBF
Save as type: *.DBF

Perform Additional Simulations

Number of Additional Simulations: 100

Single Species; Single-Season - Model Assessment - *MARK*

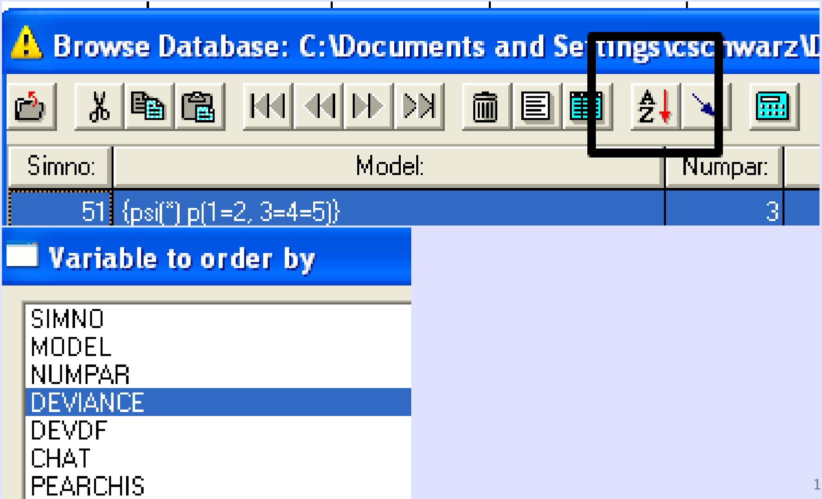
Let the program run; View the bootstrap results



Model	AICc	Delta AI		elihood	No. Pa
{psi(*) p(1=2, 3=4=5)}	163.5033	0.0		1.0000	
{psi(*) p(*)}	166.0919	2.5		0.2741	
{psi(*) p(t)}	170.3394	6.8361	0.02508	0.0328	

Single Species; Single-Season - Model Assessment - *MARK*

Let the program run; View the bootstrap results; Sort the bootstrap results



Browse Database: C:\Documents and Se tings\csc warz\D

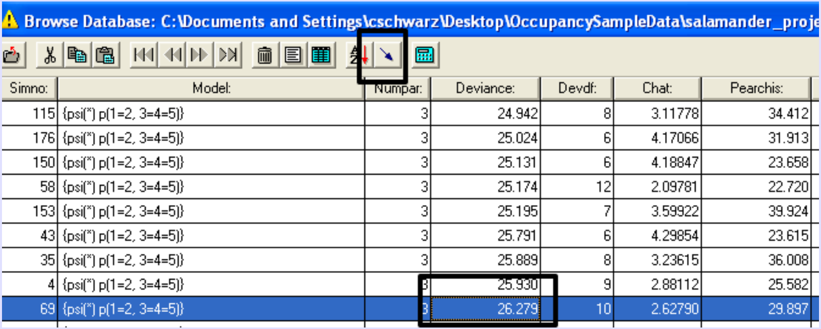
Simno:	Model:	Numpar:
51	{psi(*) p(1=2, 3=4=5)}	3

Variable to order by

- SIMNO
- MODEL
- NUMPAR
- DEVIANCE
- DEVDF
- CHAT
- PEARCHIS

Single Species; Single-Season - Model Assessment - *MARK*

Locate the observed deviation (26.11), and find how unusual is the value (click on the pointer)

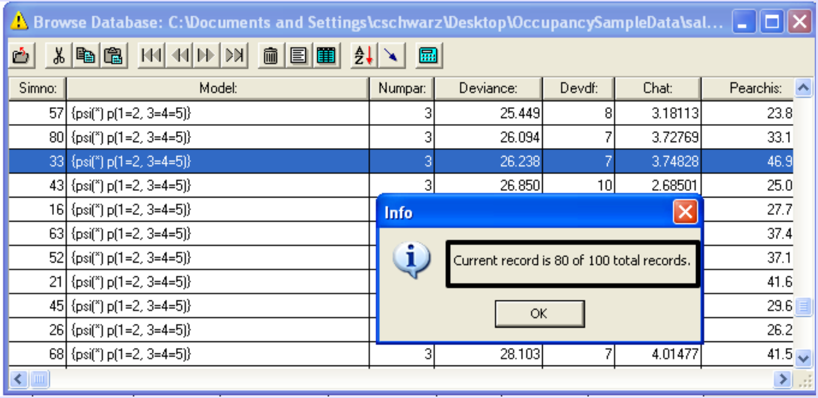


Browse Database: C:\Documents and Settings\cschwarz\Desktop\OccupancySampleData\salamander_proje

Simno:	Model:	Numpar:	Deviance:	Devdf:	Chat:	Pearchis:
115	{psi(*) p(1=2, 3=4=5)}	3	24.942	8	3.11778	34.412
176	{psi(*) p(1=2, 3=4=5)}	3	25.024	6	4.17066	31.913
150	{psi(*) p(1=2, 3=4=5)}	3	25.131	6	4.18847	23.658
58	{psi(*) p(1=2, 3=4=5)}	3	25.174	12	2.09781	22.720
153	{psi(*) p(1=2, 3=4=5)}	3	25.195	7	3.59922	39.924
43	{psi(*) p(1=2, 3=4=5)}	3	25.791	6	4.29854	23.615
35	{psi(*) p(1=2, 3=4=5)}	3	25.889	8	3.23615	36.008
4	{psi(*) p(1=2, 3=4=5)}	3	25.930	9	2.88112	25.582
69	{psi(*) p(1=2, 3=4=5)}	3	26.279	10	2.62790	29.897

Single Species; Single-Season - Model Assessment - *MARK*

How unusual is the observed deviance?



Browse Database: C:\Documents and Settings\cschwarz\Desktop\OccupancySampleData\sal...

Simno:	Model:	Numpar:	Deviance:	Devdf:	Chat:	Pearchis:
57	{psi(*) p(1=2, 3=4=5)}	3	25.449	8	3.18113	23.8
80	{psi(*) p(1=2, 3=4=5)}	3	26.094	7	3.72769	33.1
33	{psi(*) p(1=2, 3=4=5)}	3	26.238	7	3.74828	46.9
43	{psi(*) p(1=2, 3=4=5)}	3	26.850	10	2.68501	25.0
16	{psi(*) p(1=2, 3=4=5)}					27.7
63	{psi(*) p(1=2, 3=4=5)}					37.4
52	{psi(*) p(1=2, 3=4=5)}					37.1
21	{psi(*) p(1=2, 3=4=5)}					41.6
45	{psi(*) p(1=2, 3=4=5)}					29.6
26	{psi(*) p(1=2, 3=4=5)}					26.2
68	{psi(*) p(1=2, 3=4=5)}	3	28.103	7	4.01477	41.5

Info

Current record is 80 of 100 total records.

You would get an observed deviance (or larger) 20/100=20% of the time. There is no evidence of a lack-of-fit. You would be worried if the goodness-of-fit p-value is very small.

Single Species; Single-Season - Model Assessment - *MARK*

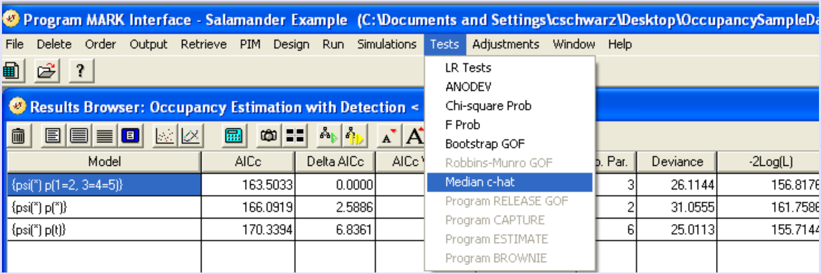
What causes lack-of-fit?

- Structural problems. You fit a $p(*)$ model when p_t varies considerably.
- Heterogeneity. Detection/Occupancy probabilities vary among sites for unknown reasons.
- Non-independence. Detection at one site "influences" detection at another site.

Try to fit more appropriate model (e.g. with covariates).
Minor modification through variance adjustment ($\widehat{c}$). Estimates may still be unbiased, but SE are multiplied by $\sqrt{\widehat{c}}$. [OK if $\widehat{c} < 5 \text{ } or \text{ } 6$.]
Estimate $\widehat{c}$ in MARK.

Single Species; Single-Season - Model Assessment - *MARK*

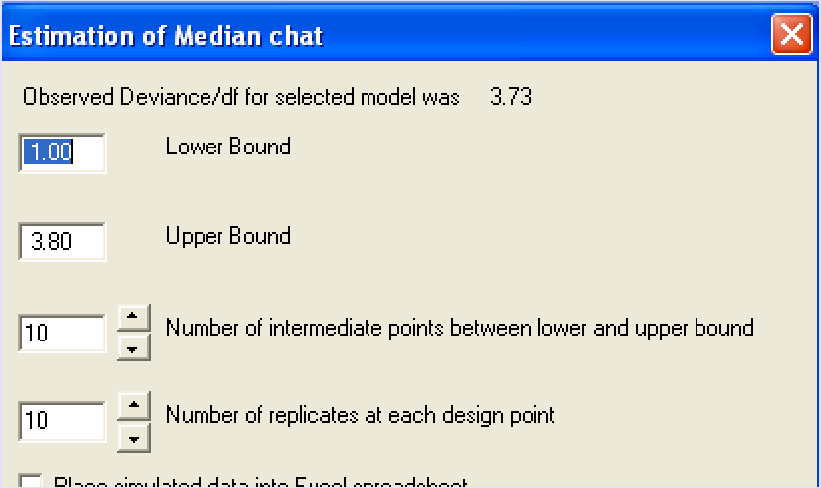
Return to results window, and select *Median C-hat* from Tests menu:



Model	AICc	Delta AICc	AICc	Par.	Deviance	-2Log(L)
{psi(*) p(1=2, 3=4=5)}	163.5033	0.0000		3	26.1144	156.817
{psi(*) p(*)}	166.0919	2.5886		2	31.0555	161.758
{psi(*) p(t)}	170.3394	6.8361		6	25.0113	155.714

Single Species; Single-Season - Model Assessment - *MARK*

Specify the number of simulations etc.



Estimation of Median chat

Observed Deviance/df for selected model was 3.73

1.00 Lower Bound

3.80 Upper Bound

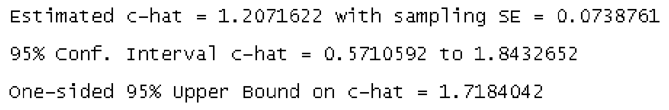
10 Number of intermediate points between lower and upper bound

10 Number of replicates at each design point

☐ Place simulated data into Excel spreadsheet

Single Species; Single-Season - Model Assessment - *MARK*

Read off the estimated value of $\widehat{c}$.

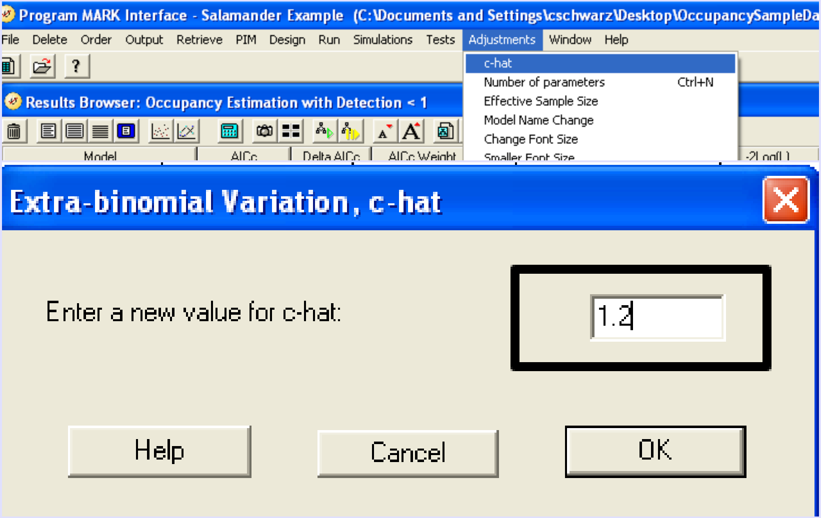


```
Estimated c-hat = 1.2071622 with sampling SE = 0.0738761

95% Conf. Interval c-hat = 0.5710592 to 1.8432652

One-sided 95% Upper Bound on c-hat = 1.7184042
```

In this case, there was no evidence of lack-of-fit so $\widehat{c}$ is 1.2 (close to 1) and could be 1. [In actual analysis, I would make no adjustment at this point.]

Single Species; Single-Season - Model Assessment - *MARK*

Enter value of $\widehat{c}$ in Results window.



Program MARK Interface - Salamander Example (C:\Documents and Settings\cschwarz\Desktop\OccupancySampleDa

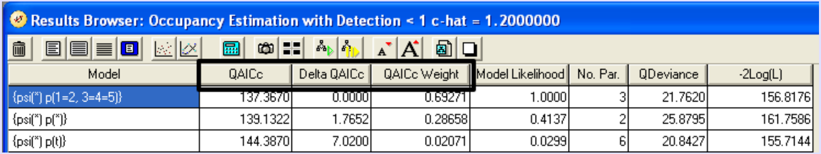
Results Browser: Occupancy Estimation with Detection < 1

Extra-binomial Variation, c-hat

Enter a new value for c-hat: 1.2

Single Species; Single-Season - Model Assessment - *MARK*

Notice that Results window now changes to QAIC etc. Interpreted analogously to AIC.

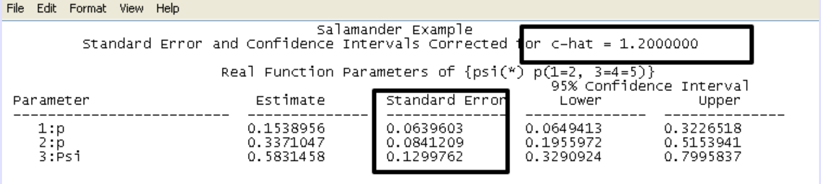


Results Browser: Occupancy Estimation with Detection < 1 c-hat = 1.2000000

Model	QAICc	Delta QAICc	QAICc Weight	Model Likelihood	No. Par.	QDeviance	-2Log(L)
{psi(*) p(1=2, 3=4=5)}	137.3670	0.0000	0.69271	1.0000	3	21.7620	156.8176
{psi(*) p(*)}	139.1322	1.7652	0.28658	0.4137	2	25.8795	161.7586
{psi(*) p(t)}	144.3870	7.0200	0.02071	0.0299	6	20.8427	155.7144

Single Species; Single-Season - Model Assessment - *MARK*

Estimates are unchanged, but SE multiplied by $\sqrt{\widehat{c}}$.



```
                              Salamander Example
       Standard Error and Confidence Intervals Corrected for c-hat = 1.2000000

                    Real Function Parameters of {psi(*) p(1=2, 3=4=5)}
                                                                 95% Confidence Interval
 Parameter                          Estimate       Standard Error      Lower           Upper
 -------------------------      --------------   --------------   --------------   --------------
    1:p                          0.1538956        0.0639603        0.0649413        0.3226518
    2:p                          0.3371047        0.0841209        0.1955972        0.5153941
    3:Psi                        0.5831458        0.1299762        0.3290924        0.7995837
```

Single Species; Single-Season - Model Assessment - *MARK*

Intuitive idea of median-$\widehat{c}$ procedure.

- In large samples deviance/df should estimate $\widehat{c}$. In our dataset is is 3.73.
- in small samples, this is biased too high.
- Generate samples of fake data (like a bootstrap) with different values of $\widehat{c}$.
- Find the value of $\widehat{c}$ where the observed deviance/df is exactly in the middle of the bootstrap values.

Single Species; Single-Season - Model Assessment - *RMark*

Unable to conduct the GOF directly with *RMark*.
You will need to export the information from *RMark* to *MARK* and do it in *MARK*.
Contact me for details.

Single Species; Single-Season - Mod Assess - Summary

- *PRESENCE* can compute gof and $\widehat{c}$ for simple models but does not adjust se's.
- *MARK* has a unified treatment of gof, $\widehat{c}$, and adjusts se.
- *RPresence* and *RMark*: Export and do in *PRESENCE* and *MARK*.
- *unmarked* can do the gof using parametric bootstrapping, but no estimate of $\widehat{c}$.
- Bayesian posterior predictive checks beyond scope of course.
- No nice visual diagnostics, but see Section 4.4.11 of MacKenzie et al. (2018).

Caught between rock and hard place here:

- With small samples, no power to detect lack of fit
- With large sample, able to detect trivial amount of lack of fit

Usually standard errors are so large that small lack of fit is unimportant.

Single Species; Single-Season - Grossbeaks

An occupancy study was made on Blue Grosbeaks (*Guiraca caerulea*) on 41 old fields planted to longleaf pines (*Pinus palustris*) in southern Georgia, USA.

Surveys were 500 m transects across each field and were completed three times during the breeding season in 2001.

Columns in the file are:

- *field* - field number
- *v1, v2, v3* - detection histories for each site on each of 3 visit during the 2001 breeding season.
- *field.size* - size of the files
- *bqi* - Enrollment in bobwhite quail initiative; does occupancy increase if field belongs to this initiative?
- *crop.hist* - crop history
- *crop1, crop2* - indicator variables for the crop history
- *count1, count2, count3* - actual counts of birds detected in each visit

Single Species; Single-Season - Exercise 02

Assess the fit of the best fitting model for the gross beak example.