USGS science for a changing world

Patuxent Wildlife Research Center

Program PRESENCE ver 12.17

Synopsis

This software was developed to enable estimation of the *Proportion of Area Occupied* (PAO), or similarly the probability that a site is occupied, by a species of interest according to the model presented by *MacKenzie et al. (2002)* [1].

Typically, species are not guaranteed to be detected even when present at a site, hence the naïve estimate of PAO given by:

```
             # sites where species detected
      PAO = --------------------------------
                 total # sites surveyed
```

will underestimate the true PAO. *MacKenzie et al. (2002)* propose that by repeated surveying of the sites, the probability of detecting the species can by estimated which then enables unbiased estimation of PAO. This model has been extended by *MacKenzie et al. (2003)*[2] that also enables estimation of colonization and local extinction probabilities. These models are briefly discussed here.

Contents

The Basic Sampling Scheme

N sites are surveyed over time where the intent is to establish the presence or absence of a species. The sites may constitute a naturally occurring sampling unit such as a discrete pond or patch of vegetation; a monitoring station; or a quadrat chosen from a predefined area of interest. The occupancy state of sites may change over time, however during the study there are periods when it is reasonable to assume that, for all sites, no changes are occurring, (e.g, within a single breeding season for migratory birds). The study therefore comprises of *K* primary sampling periods (seasons), between which changes in the occupancy state of sites may occur. Within each season, investigators use an appropriate technique to detect the species at k_j surveys of each site.

The species may or may not be detected during a survey and is not falsely detected when absent (except false-positive model). The resulting detection history for each site may be expressed as *T* vectors of 1's and 0's, indicating detection and nondetection of the species respectively. We denote the detection history for site *i* at primary sampling period *j* as $X_{i,j}$, and the complete detection history for site *i*, over all primary periods, as X_i. The single season model results when K=1, and the multiple season model for K>1.

Single Season Model

MacKenzie et al. (2002)[1] present a model for estimating the site occupancy probability (or PAO) for a target species, in situations where the species is not guaranteed to be detected even when present at a site. Let ψ be the probability a site is occupied and p[j] be the probability of detecting the species in the j^{th} survey, given it is present at the site. They use a probabilistic argument to describe the observed detection history for a site over a series of surveys. For example the probability of observing the history 1001 (denoting the species was detected in the first and fourth surveys of the site) is:

ψ x p[1](1-p[2])(1-p[3])p[4].

The probability of never detecting the species at a site (0000) would therefore be,

ψ x (1-p[1])(1-p[2])(1-p[3])(1-p[4]) + (1-ψ),

which represents the fact that either the species was there, but was never detected, or the species was genuinely absent from the site (1-ψ). By combining these probabilistic statements for all N sites, maximum likelihood estimates of the model parameters can be obtained.

The model framework of MacKenzie et al. (2002) is flexible enough to allow for missing observations: occasions when sites were not surveyed. Missing observations may result by design (it is not logistically possible to always sample all sites), or by accident (a technicians vehicle may breakdown enroute). In effect, a missing observation supplies no information about the detection or nondetection of the species, which is exactly how the model treats such values.

The model also enables parameters to be function of covariates. For example, occupancy probability may be a function of habitat, while detection probability is a function of environmental conditions such as air temperature. The model therefore allows relationships between occupancy state and site characteristics to be investigated. Covariates are entered into the model by way of the logistic model (or logit link).

A key assumption of the single season model is that all parameters are constant across sites. Failure of this creates heterogeneity. Unmodeled heterogeneity in detection probabilities will cause occupancy to be underestimated. If there is unmodelled heterogeneity in occupancy probabilities, then it is believed that the estimates will represent an average level of occupancy, provided detection probabilities are not directly related to the probability of occupancy.

Another major assumption of the MacKenzie et al. (2002) model (Single-season) is that the occupancy state of the sites does not change for the duration of the surveying. Situations where this may be violated, for instance, would be for species with large home ranges, where the species may temporarily be absent from the site during the surveying. If this process of temporary absence from the site may be viewed as a random process, (e.g., the species tosses a coin to decide whether it will be present at the site today), then this assumption may be relaxed. However, this will alter the interpretation of the model parameters ("occupancy" should be interpreted as "use" and "detection" as "in the site and detected"). More systematic mechanisms for temporary absences may be more problematic and create unknown biases. Although, users are reminded that the model assumes closure of the sites at the species level, not at the individual level, so there may be some movement of individuals to/from sites without overly affecting the model.

Model Overview

Currently, there are many types of models can be fit to detection/nondetection data within Program PRESENCE.

- Single-season models (assume sites are closed to changes in the state of occupancy for the duration of sampling)
  - Simple single-season model - detailed by MacKenzie et al. (2002)[1], estimates probability of occupancy (ψ) and probability of detection (p)
  - Spatial-dependence model - Described by Hines et al (2010) [11] which accounts for spatial correlation when surveys are chosen non-randomly.
  - Multi-method model - estimates global (ψ) and local (θ) occupancy from repeated samples of multiple methods of detection.
  - False-positive detection model - described by Miller et al. (2011) [12] which relaxes assumption that no detections occur when the species is absent and estimates the probability of a false detection.
  - Multi-state model - uses additional codes in the detection history data to denote the state of the species in a particular site/survey (0=not detected, 1=detected in state 1, 2=detected in state 2) and estimates the probability of sites being in each state.
  - Two-species model - computes occupancy and detection probabilities with interactions when there are two species present.
  - Heterogeneity(Royle/Nichols) model - described by Royle and Nichols (2003), estimates occupancy and abundance from *count data*, assuming detection depends on species abundance. [15].

- Staggered-entry model - relaxes closure assumption such that a site may -locally colonize and go locally extinct once during the survey period (ie., delayed arrival and/or early departure).
    - Royle-point-count model - described by J.A. Royle et al. (2004) [10] which uses species counts instead of presence/absence for single-season data.
- Multi-Season or Dynamic models:
    - Standard multi-season model - multiple season extension detailed by MacKenzie et al., allows estimates of local colonization and extinction of sites between seasons. (2003)[2]
    - False-positive detection model - described by Miller et al. (2013) [17] which relaxes assumption that no detections occur when the species is absent and estimates the probability of a false detection.
    - Staggered-entry model - relaxes closure assumption such that a site may -locally colonize and go locally extinct once during the surveys (ie., delayed arrival and/or early departure).
    - Heterogeneous-detections model - multiple season 'mixture' model, which assumes sites belong to one of two unidentifiable groups (with different detection probabilities).
    - Multi-state models - extends the multi-season model to allow for multiple occupancy states (eg., 0=not occupied, 1=occupied, 2=occupied with breeders) and estimates probabilities of being in each state as well as probabilities of transitioning between states.
    - Integrated-habitat-occupancy models - models changes in occupancy state in relation to changes in habitat state.
    - Two-species model - computes occupancy, colonization, extinction and detection probabilities with interactions when there two species present.

In all models, estimated parameters (ψ, p, γ, ε, ...) may be modelled as functions of site-specific, or site-survey-specific covariates.

Downloading Program PRESENCE

- Program PRESENCE is available at: https://www.mbr-pwrc.usgs.gov/software/presence.shtml#download.
- Although PRESENCE is a Windows©-based program, it has been sucessfully run on Linux and Mac computers (using Wine or Parallels).
- (Click here to go to download PRESENCE page)

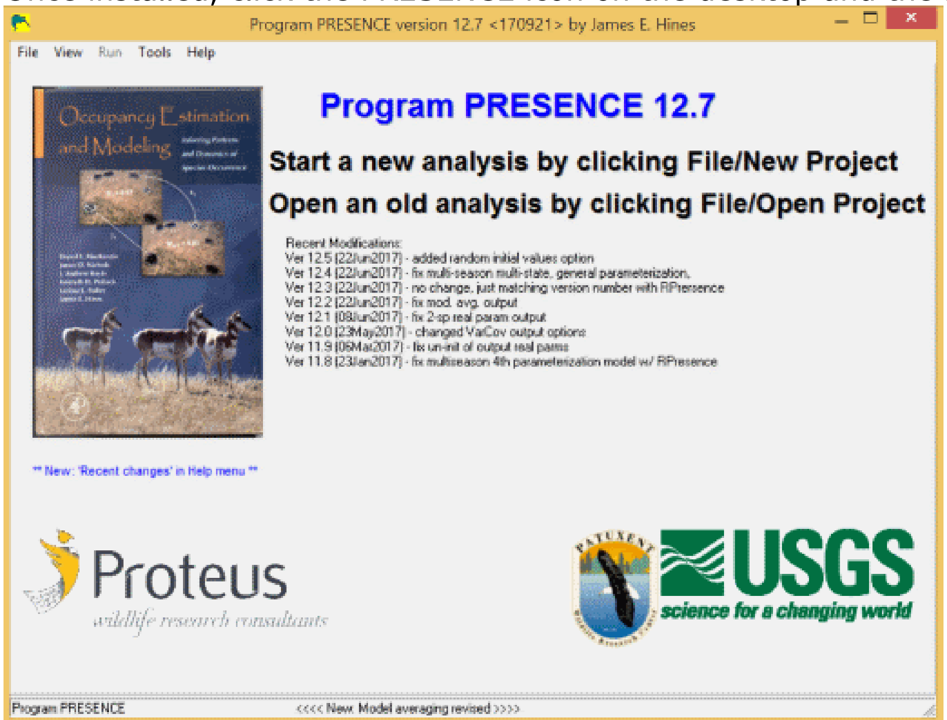
Installing Program PRESENCE

Here is the procedure to install on a Windows© system:

- Download and run the installation program (setup_presence.exe).
- Suggestion: If you are unable to install the program due to access restrictions, install PRESENCE in a different folder (not "C:\Program Files"). You can create a folder named "C:\Pgms" to hold the program which will not be write-protected (like C:\Program Files). If you are unable to run the setup program due to lack of administrative rights, download the PRESENCE zipfile (PRESENCE (no install/setup program)) from the website and unzip the contents to a folder that you have write-access to.
- Note: From time to time, it's a good idea to check to make sure you have the latest version of PRESENCE. This might prevent the case of reporting a bug which has already been fixed.

Running the program

- Once installed, click the PRESENCE icon on the desktop and the following screen appears:



- There are several "tutorials" in the *Help* menu, showing step-by-step instructions on running some of the models with sample data.
- Input data and model output files are stored in a "project folder". To start an analysis, go to the "File" menu and select "New Project".
- There are a couple of options for inputting data to the program:
  - Use copy/paste functions to copy data from your spreadsheet program into PRESENCE. This is the most common option.
  - Create a "comma-separated-value" (csv) files with detection and covariate data.

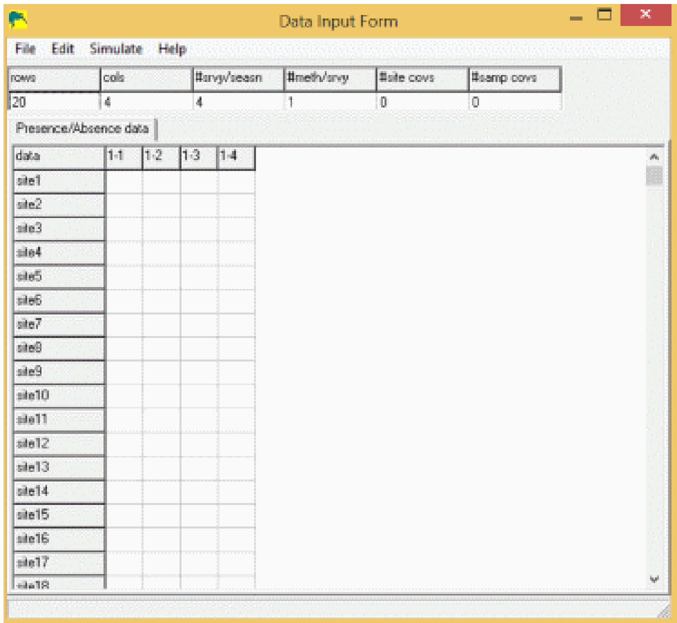
The 1st option above is the recommended choice since most spreadsheet programs will automatically backup the data as it is entered. In case the power fails, you might be able to retrieve previously entered data.

Once the data are entered into the program, it must be saved (use menu option File/Save) in order to build and run models. The saved file will have an extension (last 4 characters of filename) of ".pao". This file will contain both presence/absence data and site and sample covariate data.

To build and run models, a 'project-folder' is created by the program. This file will contain the results of each model in separate files, as well as the input (pao) file and summary results (pa3) file.

To start a new analysis, Select 'File/New project' from the menus. A form will appear which will hold the information about the analysis, including title, filenames, data-type, and numbers of sites/occasions/covariates. At this point, you may use a previously-created input file (something.pao) by clicking the 'select file' button, or go to the input screen by clicking the 'Input form' button. Clicking the 'select file' button allows you to navigate to the folder containing the input file and select the file. Clicking the 'Input Form' button displays a new form with a tabbed spreadsheet-like interface.

Input Form



To enter data into this form, click on the first element (site 1, sample #1), and enter '1' (without quotes) if the species was detected at site 1, sample #1, '0' if the species was not detected, or '-' if this site was not sampled. The 'Tab' key will move the cursor to the next sample (or use the mouse) where you can enter the data for site 1, sample 2. Since most users will have data prepared in some other form (eg., spreadsheet or database), this entry method would (should) never be used as it introduces another source of error in the data.

If your data is already entered in a spreadsheet program, you can open that program, select all the site/sample data (no headers or other fields), and click 'Edit/Copy' from the menus. Then, go back to PRESENCE, and click the 'Edit/Paste values' from the menu. If your spreadsheet contains sitenames in the first column, you can include these in the selection-edit-copy, then select 'Edit/Paste w/sitenames' in the PRESENCE menu.

If you have covariate data. (e.g., weather or effort data), you can enter these by changing the number of covariates in the appropriate box at the top of the form, then clicking the appropriate tab and entering data as was done with the presence/absence data.

Once the data are entered (or simulated), click 'File/Save' from the menu, then click 'File/Close'. *This is an important step, as PRESENCE will not be able to use data in the form unless it has been saved.*

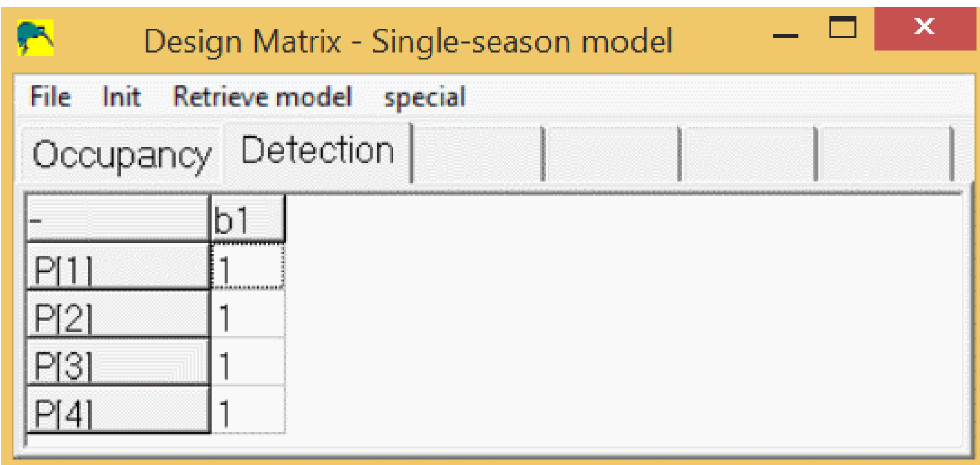
Next, click the 'select file' button on the 'enter specifications' form, and use the Windows file selector to navigate to the folder where the saved file resides and select the file. If you've used the input data form and saved the file from the file menu, the input filename and results filename will already be entered for you.

Once you've entered the input (pao) and results (pa3) filenames, click the 'OK' button to create the PRESENCE project folder. This will close the specification window and open a 'Results summary' window. You are now ready to compute estimates under pre-defined models, or build your own 'Custom' model.

Simple single-season model

Program PRESENCE allows you to build models where the model parameters (occupancy, detection, colonization, extinction,...) depend on various covariate information collected in conjunction with the detection data. To allow flexibility, models are defined by using a 'design-matrix'. Although the mention of "design-matrices" may cause anxiety in non-statistician biologists, they can be thought of as a list of simple mathematical equations relating quantities to estimate (beta parameters) to real parameters of interest (real parameters, ψ, p, γ, ε).

In the design-matrix, the beta-parameters are represented by the columns, and the real parameters are represented by the rows. As an example, let's look at building a simple single-season model where occupancy is constant across all sites and detection is constant across all sites and surveys. When you click 'Run/Analysis-single-season/simple-single-season', a design-matrix form appears. This form contains a tab for the occupancy model parameter, ψ (psi), and a tab for the detection model survey-specific parameters (p1,p2,p3,p4). If you click the 'occupancy' tab, there will be a spreadsheet with 1 row (psi) and 1 column (a1). The estimated ("beta") parameter is 'a1' and the real model parameter is psi, which is computed as psi = 1 * a1. If you click the "Detection" tab, a spreadsheet will appear which contains 4 rows and 1 column. By default, the "real" model parameter p1, will be computed as: p1 = 1 * b1. The real model parameters p2, p3, and p4 will also be computed as: p2 = 1 * b1, p3 = 1 * b1, and p4 = 1 * b1. So, the program will estimate 1 parameter, a1, which will yield a value for psi, and another parameter, b1, which will yield the same value for p1, p2, p3, and p4. Here is the design matrix for detection with constant detection (p1=p2=p3=p4):



Another simple single-season model

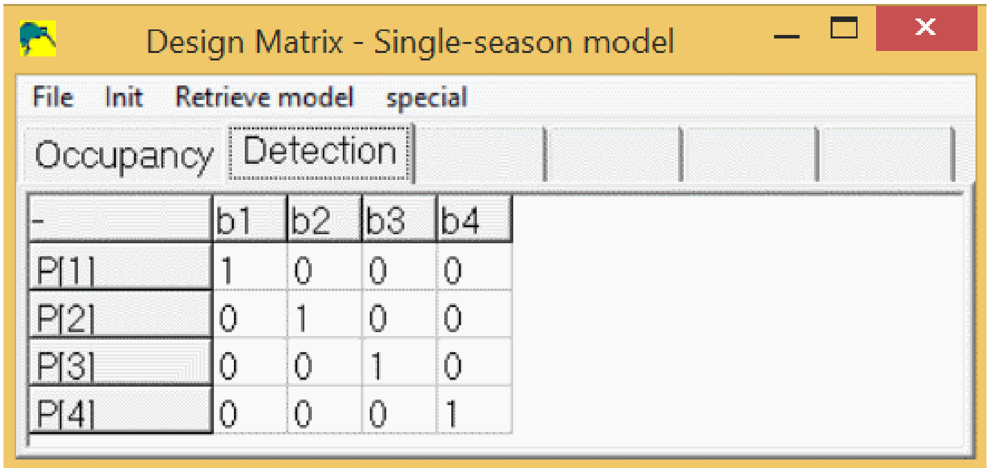
To build a model where detection probability is survey-specific (different for each survey), we need different beta parameters for each detection probability, p1, p2, p3, and p4 (ie.,

$p1 = 1 * b1$
$p2 = 1 * b2$
$p3 = 1 * b3$
$p4 = 1 * b4$

). We can write those 4 equations for p1-p4 such that they include b1-b4 as:

$p1 = 1 * b1 + 0 * b2 + 0 * b3 + 0 * b4$
$p2 = 0 * b1 + 1 * b2 + 0 * b3 + 0 * b4$
$p3 = 0 * b1 + 0 * b2 + 1 * b3 + 0 * b4$
$p4 = 0 * b1 + 0 * b2 + 0 * b3 + 1 * b4$

This coefficients of the beta parameters (b1-b4) yields the detection design matrix:



This design-matrix (zeros, with 1's on diagonal) is a special matrix called the "identity" design matrix. It gives a one-to-one correspondence between the estimated "beta" parameters (b1-b4) and the real parameters (p1-p4). In program PRESENCE, this can be achieved by the menus, "Init/Full Identity".

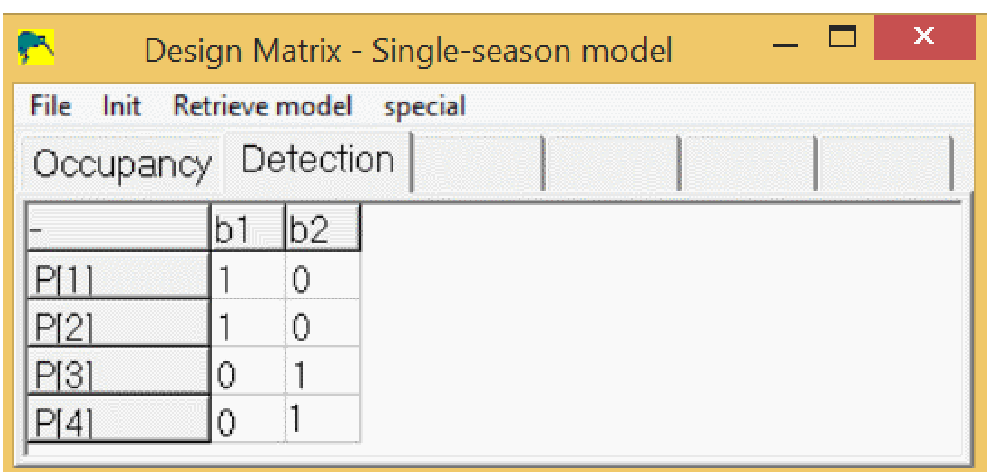
We now have 5 estimated "beta" parameters (a1,b1,b2,b3,b4), which will be used to compute 5 model "real" parameters (psi,p1,p2,p3,p4). .

A 3rd simple single-season model

Suppose there was something different about the first two surveys from the second two surveys (eg, sunny for surveys 1-2, rainy for days 3-4). In this case, detection probabilities, p(i), are not constant as in the first model, and not different each sample, as in the second model. Here, we would like to build a model where the first two detection probabilities are the same, but different from the last two detection probabilities. In this case, we would need to estimate 1 parameter for the first two p's, and another parameter for the second two p's (and a third parameter for psi). Here is how we would want the p's computed:

p1 = *1* * b1 + *0* * b2
p2 = *1* * b1 + *0* * b2
p3 = *0* * b1 + *1* * b2
p4 = *0* * b1 + *1* * b2

So, the detection spreadsheet would contain 4 rows (p1-p4), and 2 columns (b1,b2) and would look like this:



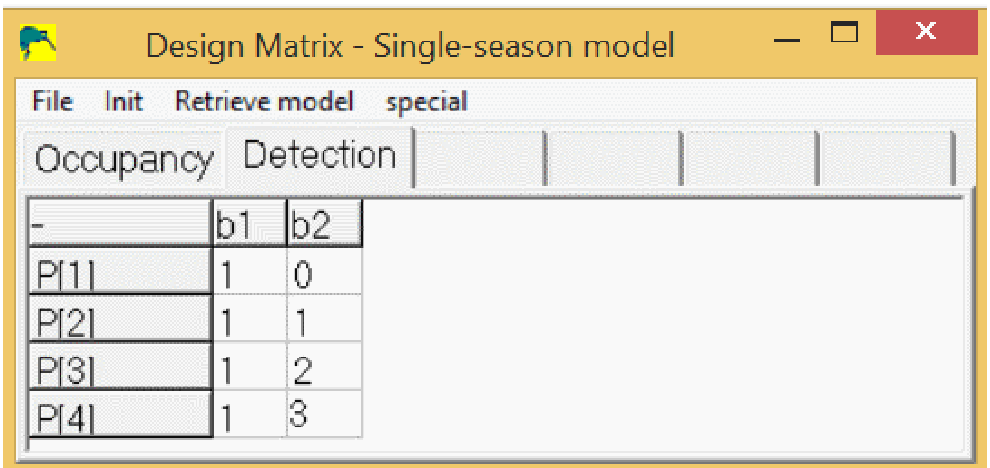
To modify the design matrix in PRESENCE to look like this, right-click on a cell in the design matrix, then select "add cols" or "del cols" to get the desired number of columns, then click in each cell to change it's value to 0 or 1 to match the one above.

A 4^{th} Model

In this example, suppose detection probabilities, p(i), are hypothesized to be increasing by a constant amount over the surveys. So, the second detection probability would be equal to the first detection probability plus a constant (X), and the third would be equal to the first + 2*X, and the last detection probability would be equal to the first + 3*X. We would need to estimate two quantities: detection in survey 1 (p1), and the rate of increase/decrease in detection from one survey to the next. We can use b1 for detection in survey 1 (p1=b1), and b2 for the rate of change (X) in detection. Here is how to write the formulae for the p's:

$p1 = 1 * b1 + 0 * b2$
$p2 = 1 * b1 + 1 * b2$
$p3 = 1 * b1 + 2 * b2$
$p4 = 1 * b1 + 3 * b2$

So, the detection spreadsheet would contain 4 rows, and 2 columns (b1,b2) and would look like this:



-	b1	b2
P[1]	1	0
P[2]	1	1
P[3]	1	2
P[4]	1	3

A 5th Model (with covariates)

In the previous examples, detection probabilities were assumed to be the same for each site. By using covariates, we can compute a different detection probability for each site (or possibly each site/sample combination). Without covariates, this assumption would cause us to estimate a large number of parameters (20 for the first pre-defined model, and 80 for the second). By using covariates, we can compute p as:

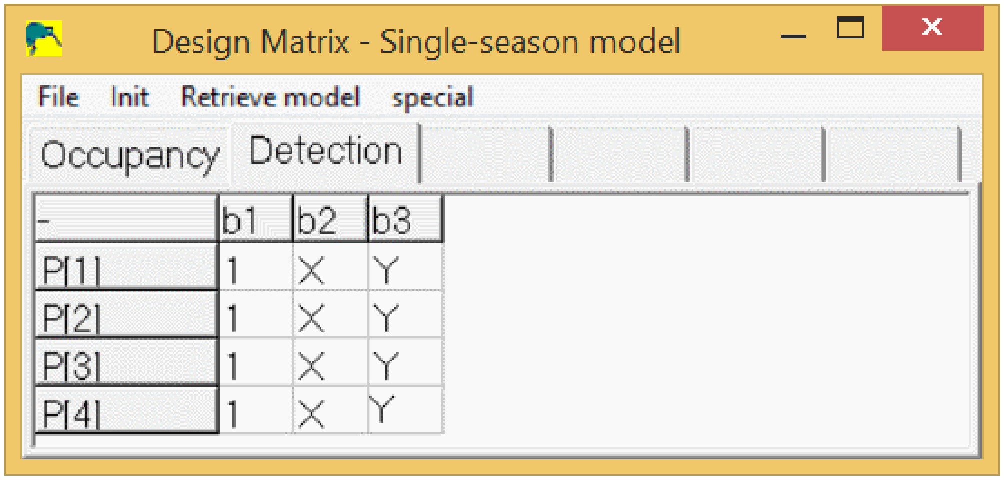
$p(\text{site } i, \text{survey } j) = 1 * b1 + X(i,j) * b2$

where X(i,j) is the value of the sample-covariate at site i, sample j. Here, p(i,j) is equal to a base detection probability, (intercept), (1*b1) plus an effect (b2) of the covariate, X(i,j). If b2=0 then there is no effect of the covariate (p=constant). If b2>0 then there is a positive effect of the covariate (higher covariates yield higher p's), and if b2<0 then there is a negative effect of the covariate (higher covariates yield lower p's).

If you had two covariates for each site/sample, you could compute p as:

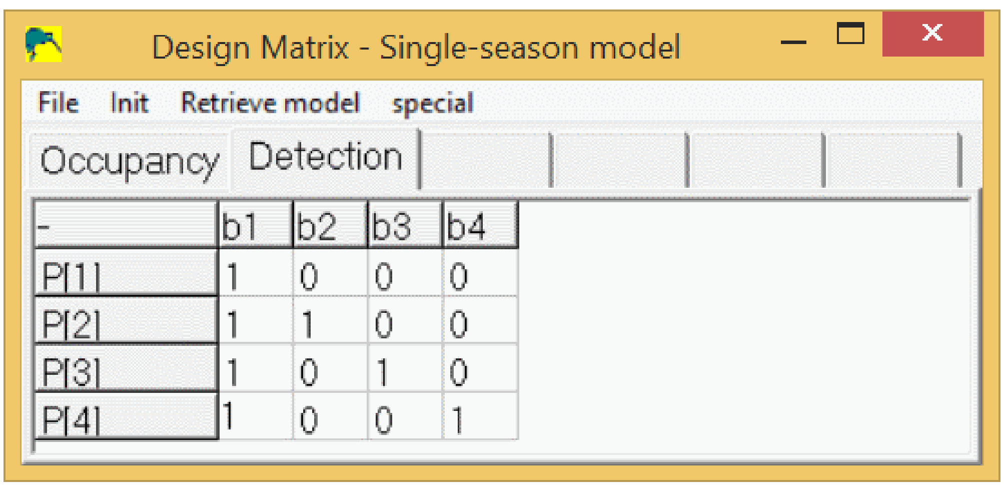
$p(\text{site } i, \text{survey } j) = 1 * b1 + X(i,j) * b2 + Y(i,j) * b3$

where Y(i,j) is the value of the 2^{nd} covariate at site i, sample j. The detection design matrix for the case with 2 covariates would look like this:



To modify the design matrix in PRESENCE to look like this, right-click on a cell in the design matrix, then select "add cols" or "del cols" to get the desired number of columns, then click in the first cell under b1 and select "Init/Constant" from the menus. Then, instead of entering "X" and "Y" in the other cells, use the menus "Init/Individ.Covariates/*X" after clicking on the 1st cell under "b2" and the menus "Init/Individ.Covariates/*Y" after clicking in the 1st cell under b3. You could manually type "X" and "Y" in the appropriate cells, but PRESENCE will not forgive you if you make a mistake in typing the covariate name (including using the wrong case).

This method can be applied to any of the parameters (ψ, p, γ, ε, θ,...) in the models. Also, the saying, "*There is more than one way to skin a cat.*" applies (disclaimer: no cats were harmed in the creation of this document.), meaning the following design matrix will yield the same log-likelihood, AIC value, and real parameter estimates as the 2nd model described above:



This design matrix might look more familiar to the more statistically inclined and can be interpreted as follows:

- b1 = intercept term, or detection probability in survey 1,
- b2 = effect of survey 2, or difference in detection probability in survey 2 versus survey 1,
- b3 = effect of survey 3, or difference in detection probability in survey 3 versus survey 1,
- b4 = effect of survey 4, or difference in detection probability in survey 4 versus survey 1,

A First Example

Now seems like a good time to run through an example. Start program PRESENCE and select 'File/New Project' from the menus. When the 'Enter Specifications' form appears, click the 'Input Data Form' button.

The input data form will contain only 1 tab, for the presence/absence data. We're going to simulate some data in this form. Let's assume we're dealing with a species which has an occupancy rate (ψ) of 0.60 (60% of areas contain at least 1 individual of the species). Also, assume/pretend detection probability is lousy in the beginning, p(1)=0.2, and gets better on each successive sample, p(2)=.4, p(3)=.6, p(4)=.8. This is enough information to generate data for the single-season data-type. To generate presence/absence data, select 'Generate data' from the 'Simulate' menu and enter ψ (0.6) when prompted. Next, enter '0.2,0.4,0.6,0.8' when prompted for the detection probabilities (p) and the table will clear and be filled with randomly generated presence/absence data with those parameters. Click 'File/Save as' and save the file with the name, 'simdata1.pao'. When asked about using the last column for the frequency, click 'No', and enter something for a title. Then, click 'File/Close'.

Next, click the 'Click to select file' button and select the simulated data file we just created (simdata1.pao). The program will fill in the boxes for the filename and results filename. Enter 'simulated data w/ psi=.6, p=.2,.4,.6,.8' in the title box, and click the 'OK' button.

You should see an empty results browser at this point. To run the first pre-defined model, click 'Run/Analysis:single season' from the menus. For the pre-defined models, the program automatically fills in the model name. This name can be anything, but it's best to make it something easily recognized. (I'll describe a common convention for naming later.) In the 'Model' box, you'll see that 'pre-defined' is already selected, and 6 pre-defined models are listed (with the first one selected). Let's start with this model, but first check the 'list data' option. Click 'list-data', then click the 'OK to run' button.

Once you click the 'OK to run' button, you might see another window flash by (perhaps not if you have a fast computer), then a dialog box appears with a short summary of the results of that model. Click 'Yes' to include the output of that model in the results browser. (You might click 'No' in the case where you accidentally run a model which was previously run.) After clicking 'Yes', the summary information from that model is displayed in the results browser.

To view the estimates of psi, and p from this model, use the mouse to position the cursor over the name of the model, '1 group, Constant P', then click with the right mouse button. A pop-up menu will appear. Position the mouse over 'View model output' and click with the left mouse button. This will cause a Notepad window to appear with the results. Look at the output and note the estimates of ψ(Psi), and p. (I got .7267, and .4300, but yours will be different.)

Next, let's run another pre-defined model - one with survey-specific p's. Close the notepad window with the results, then click 'Run/Analysis:single season' from the menus. Click '1 group, survey-specific p' in the 'Model' box (note model name changed for you), then click 'OK to run'. Click 'Yes' to include the results of this model in the results browser, then position the mouse over the model name, right-click, then left-click 'view model output'. In the notepad window, note ψ (Psi), and p(1),p(1),p(3) and p(4). With such a small number of sites and surveys, estimates may be very different from the true values (ψ=.6 , p=.2,.4,.6,.8).

Spatial Dependence in Single-season Model

Replication of surveys may be temporal or spatial in nature. That is, the same sites can be sampled at different points in time, or a site may be broken into multiple locations and sampled once. A common example of spatial replication is transect sampling, where observers walk along a trail and attempt to detect sign of the species (eg., bird singing, mammal scat, visual spotting) at specific points along the transect. A common problem with spatial replication is that an individual of a wide-ranging species (eg., tigers) may roam among several spatial replicate stops. So, if one point along a transect has an individual nearby, detection probability will be higher than points which are not near an individual. This leads to a form of detection heterogeneity where there is autocorrelation in sample detection probabilities. The same sort of autocorrelation can occur in aural surveys where a 10-minute detection survey is broken into intervals of 3,3 and 4 minute detection intervals. The "Single-season correlated detections" model addresses this problem. Surveys are conducted along trails such that when a species is found at one sample, nearby samples have a much higher (or lower) probability of the species being present than those farther away. This can be accounted for by adding two new parameters, θ, θ'. θ is the probability that the species is present locally, given the species was *not present* in the previous sample, but globally present at some point along the transect. θ' is the probability that a species is present locally, given it *was locally present* at the previous sample and globally present along the transect.

An example detection history might be:

01011

Here, the species was detected at the 2nd, 4th and last samples (segments of transect line), but not detected at the 1st and 3rd samples. The probability of this history would be represented by:

$\psi\{(1-\pi)\ [(1-\theta_1)\theta_2+\theta_1(1-p_1)\theta'_2]\ +\pi\ [(1-\theta'_1)\theta_2+\theta'_1(1-p_1)\theta'_2]\ \}\ p_2[(1-\theta'_3)\theta_4+\ \theta'_3(1-p_3)\theta'_4]\ p_4\theta'_5p_5$

In this model, we've added a new parameter, π, which is the proportion of sites which are locally occupied *before* the first survey. If all transects begin at a boundary, such that it is impossible for the species to be locally present before the first survey (eg., start at a lake, or road), then this parameter may be fixed to zero. Another option for the π parameter is to assume that it is equal to some sort of average of θ and θ'. This can be achieved in PRESENCE by fixing π to "EQ", which is an abbreviation for "equalibrium", meaning the value of local occupancy obtained if a random survey is chosen from all surveys.

When this model is chosen, the θ parameters will appear in the design matrix window in the same tab as the occupancy (ψ) parameter. This model is described in Hines (2010)[11]

Note: this model allows the theta's to be segment-specific. This may be possible if the detection parameters (p) are constant among segments (or a function of a survey-specific covariate). This model isn't identifiable (no single solution for parameters) if both θ's and detection probabilities are allowed to be segment-specific.

This model is a special case of new multi-season Spatial Dependence model where there is only 1 season. So, if you have single-season data (number of surveys per season = total number of surveys), just run the multi-season correlated detections model.

Single-season False-positive detection model

The False-positive detection model (Miller et. al. (2011). [12]) extends the single season model by relaxing the assumption that detections do not occur when the species is not present. This type of model might be used when one suspects that it is likely that detections are not being recorded perfectly. For example, a detection for species A might actually be a similar species, but recorded for species A. A common example of the need for this model is the case where some observers are 'untrained' or 'inexperienced' and others are "experts". A requirement for this model is that some observations must "known", so the probability of a false detection can be estimates. For example, in aural surveys, where some detections are based on auditory information (where there is a possibility of mistaking the call of another species for the target species) and some detections are based on visual information (where the observer is sure the species is present). Both types of data are required in order to obtain estimates for this model.

Parameters:

- ψ - probability that the area is occupied by the species,
- $p11_i$ - probability that the species is detected, given the site is occupied.
- $p10_i$ - probability that the species is detected, given the site is **un**occupied.
- b_i - probability of an 'assured' detection, given the site is occupied and a detection occured.

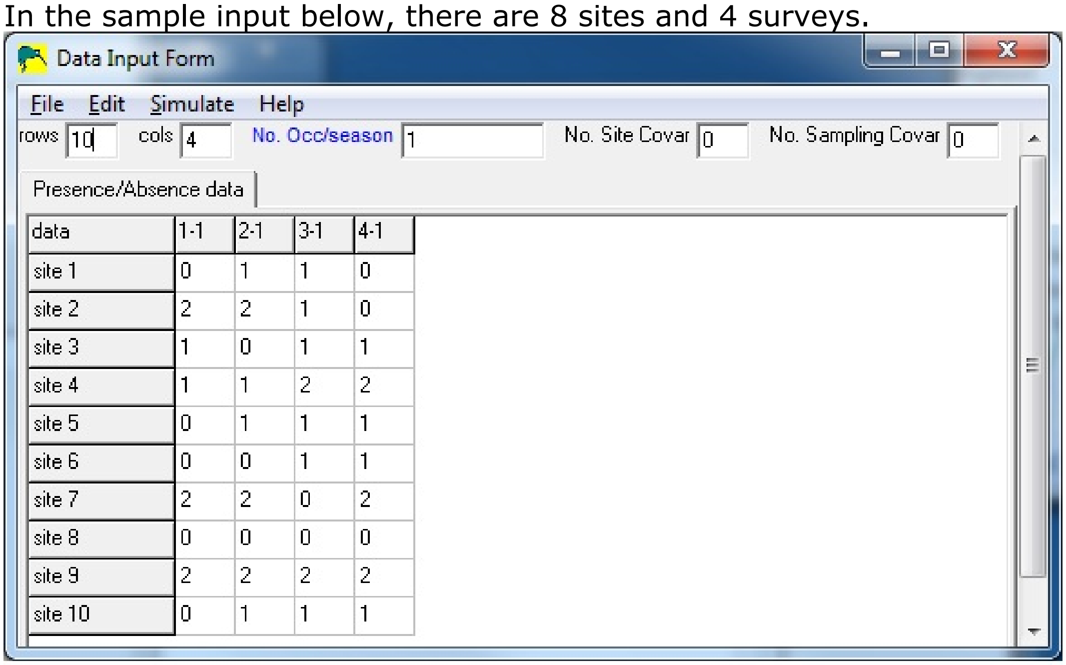
An 'assured' detection is one where there is no doubt about whether the species is present.

Data input for this model is as follows:

site1 h_1 h_2 h_3...
site2 h_1 h_2 h_3...
site3 h_1 h_2 h_3...
:
:

where
h_j = 2 if detection at site for survey j, and detection is certain;
h_j = 1 if detection at site for survey j, but detection is uncertain;
h_j = 0 if no detection

In the sample input below, there are 8 sites and 4 surveys.



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

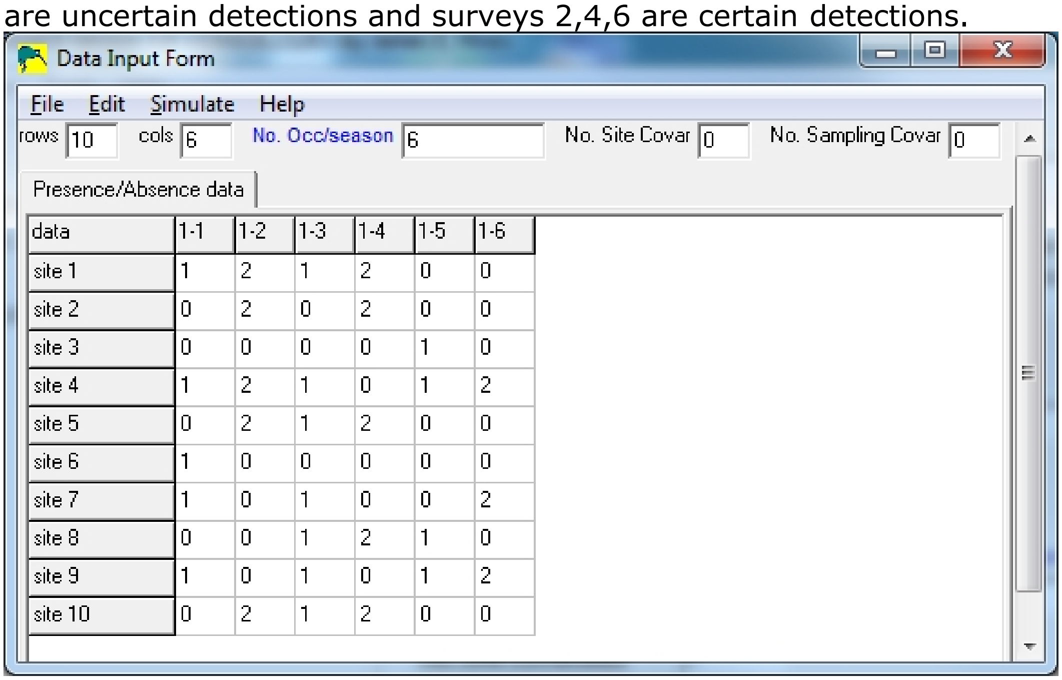
In this example, some detections were uncertain (auditory detection of species) and some detections were certain (eg., visual or both auditory and visual detection of species). Site 1 contains an uncertain detection in survey 3, and site 2 contains an uncertain detecion in surveys 1 and 3, and a certain detection in survey 2.

Alternative False-positive sampling scheme

Instead of having surveys with combined uncertain and certain detections, the sampling scheme may involve two methods of detection, where one method yields certain detections (eg., DNA sample, visual detection), and one method yields uncertain detections (eg., auditory detection). The False-positive model can still be used for these data. The input data would have to be entered such that each survey would occupy two columns in the input file: one column for the uncertain detections and another for the certain detections. Uncertain detections are coded as '1' and certain detections are coded as '2' (as above). The 'trick' needed to make this model work is to fix all of the 'b' parameters (since the probability of a detection being 'certain' is known for all detections). For the columns which correspond to uncertain surveys, the 'b' parameter should be fixed to 0, and the 'b' parameter for columns which correspond to certain detections should be fixed to 1. The probability of false detections, 'p10', should be fixed to 0 for columns which correspond to 'certain' detection surveys.

For example, if there are 3 surveys with 'certain' and 'uncertain' detections in each survey, the input detection-history data would contain 6 columns. Columns 1,3,5 would contain uncertain detections in each of the 3 surveys (0=not detected, 1=detected). Columns 2,4,6 would contain certain detections (0=not detected, 2=detected). As far as PRESENCE is concerned, there are 6 surveys, but you will know that there are really only 3 surveys, with two methods per survey.

In the sample below, there are 10 sites and 6 surveys, with the implication that surveys 1,3,5 are uncertain detections and surveys 2,4,6 are certain detections.


Data Input Form

File Edit Simulate Help

rows 10 cols 6 No. Occ/season 6 No. Site Covar 0 No. Sampling Covar 0

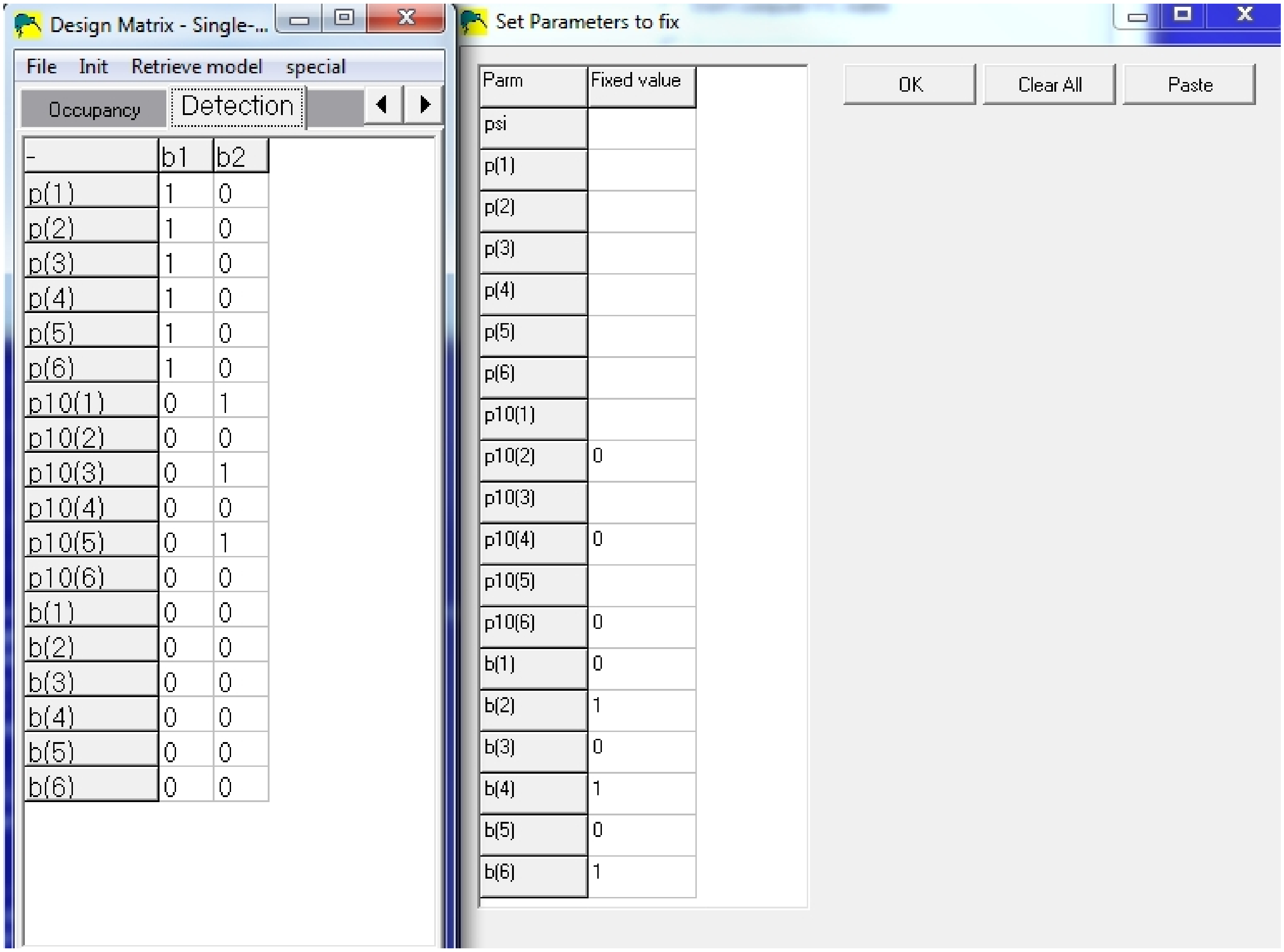
Presence/Absence data

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

Notice that columns 1,3,5 only contain 0's and 1's, and columns 2,4,6 only contain 0's and 2's.

When building the model, the detection design matrix will now contain 6 detection probabilities (p11), 6 false-positive probabilites (p10), and 6 'certainty probabilities' (b). Since there cannot be any false-positive detections in surveys 2,4,6, those p10's need to be fixed to zero (p10(2)=0, p10(4)=0, p10(6)=0). Also, since there are no certain detections in surveys 1,3,5, those b's need to be fixed to 0 (b(1)=0, b(3)=0, b(5)=0).

The design matrix and fixed-parameters windows would look like this:



Notice that any parameter which is fixed (right window) contains all zeros for that row in the design matrix window (left window). This example design matrix shows a model where detection probabilities are the same for both methods (p(odd)=p(even)), which is probably not very realistic. If the methods are drastically different, one would expect the detection probabilites to be different and model the p's differently.

Staggered-entry model

This model relaxes closure assumption such that a site may locally colonize and go locally extinct once during the surveys (ie., delayed arrival and/or early departure). It is descirbed by Kendall et. al 2013.[13]. Input data is of the same format as the single-season model.

Parameters:

- ψ - probability that the area is occupied by the species,
- b_i - probability of species entering between surveys i and i+1, given not entered yet,
- d_i - probability of species departing between surveys i and i+1, given presence,
- p_i - probability of detecting species in survey i, given presence.

Royle/Nichols Occupancy Model for Abundance-induced Heterogeneity

The Royle/Nichols heterogeneity model (Royle and Nichols, 2003) [7] estimates population size from temporally replicated presence/absence data (from point-counts or other types of surveys) at a number of sample sites. This model assumes that heterogeneity in detection probability among sites is due to heterogeneity in abundance (more individuals lead to higher probability of detecting the species at the site). Input data for this model are the presence/absence (1/0) of the species at each survey at each sample site.

Parameters estimated under the assumption of a Poisson distribution:

- λ - population density (per site),
- r - probability of detection (per individual of the species)

The species-level probability of detection can be computed using r and λ as:

$p_{i,j} = 1 - (1-r_{i,j})^{\lambda_i}$

Two-Species Model

The two-species model (MacKenzie et al., 2004) [16] extends the single season model in another way by allowing the computation of occupancy parameters of two species along with conditional probabilities of occupancy when the other species is present or detected.

Parameters:

- ψ^A - probability that the area is occupied by species *A*,
- ψ^B - probability that the area is occupied by species *B*,
- φ - species co-occurance ($\psi^{AB}/(\psi^A*\psi^B)$), {$\psi^{AB}$=probability that area is occupied by both species},
- p^A - probability of detecting species *A*, given species *B* is not present,
- p^B - probability of detecting species *B*, given species *A* is not present,
- r^A - probability of detecting species *A*, given both are present,
- r^B - probability of detecting species *B*, given both are present,
- λ - species co-detection ($r^{AB}/(r^A*r^B)$), {r^{AB}=probability of detecting both species, given both are present},

Input data for this model is in the same form as the single-species, single-season model except that the first half of the detection history records are assumed to be species A, and the second half of the records are assumed to be species B. So, if there are 60 sites, the input would consist of 120 detection history records. Records 1-60 would be the site-detection history records for sites 1-60, species A, and records 61-120 would be the site-detection history records for sites 1-60, species B.

Alternatively, data could be coded without doubling the number of sites. In this case, detections would consist of the following codes:

- 0 = no detection of either species
- 1 = detection of species A only
- 2 = detection of species B only
- 3 = detection of both species

Alternate parameterization

Since two of the parameters in the default parameterization are not probabilities bounded by the interval (0 - 1), numerical problems can arise. (eg., if ψ^A is zero, φ would be undefined.)

An alternate parameterization was developed, using conditional probabilities as parameters, which is more numerically stable. The parameters are:

- ψ^A - probability that the site is occupied by species *A*,
- ψ^{BA} - probability that the site is occupied by species *B*, given species A is present
- ψ^{Ba} - probability that the site is occupied by species *B*, given species A is not present
- p^A - probability of detecting species *A*, given only species *A* is present,
- p^B - probability of detecting species *B*, given only species *B* is present,
- r^A - probability of detecting species *A*, given both are present,
- r^{BA} - probability of detecting species *B*, given both are present, and species A was detected
- r^{Ba} - probability of detecting species *B*, given both are present, and species A was not detected

Using this parameterization, quantities from the other parametrization can be computed. (eg.,

$\psi^B = \psi^A*\psi^{BA}+(1-\psi^A)*\psi^{Ba}$
$\varphi = \psi^A*\psi^{BA}/(\psi^A*\psi^B)$

Two-Species Model with False-positive detections

The Single-season,two-species model (MacKenzie et al., 2004) is extended to allow mis-identification of the species (see Chambert et. al. 2018[18]).

Additional parameters:

- ω^A - ((oA) in PRESENCE) probability of erroneously detecting species B at a site where only species A is present and species A has also been correctly detected at that site-occasion).
- ω^a - ((oa) in PRESENCE) probability of erroneously detecting species B at a site where only species A is present and species A has *not* been correctly detected at that site-occasion).
- ω^B - ((oB) in PRESENCE) probability of erroneously detecting species A at a site where only species B is present and species B has also been correctly detected at that site-occasion).
- ω^b - ((ob) in PRESENCE) probability of erroneously detecting species A at a site where only species B is present and species B has *not* been correctly detected at that site-occasion).
- η^A - ((cA) in PRESENCE) probability that only species A is confirmed at a site that was actually occupied by both species (e.g., only specimens of species A were collected and sent for genetic analyses).
- η^B - ((cB) in PRESENCE) probability that only species B is confirmed at a site that was actually occupied by both species.

Additional data: In order to obtain estimates for this model, additional data are required, in addition to the standard two-species detection data. This additional data consists of a code for each site and survey, indicating which species were confirmed at the site-survey. So, these data should be in the same format as the detection data. For example:

```
Detection data:                               Confirmation data:
survey  1    2    3    4    5    6            1    2    3    4    5    6

[1,]    2    2    0    2    0    2            0    2    0    2    0    2
[2,]    0    0    0    3    0    2            0    0    0    0    0    0
[3,]    2    0    0    2    0    0            1    0    0    0    0    0
[4,]    3    2    2    2    0    2            3    0    2    0    0    0
```

Explanation:

- At site, [1,], only species 2 was detected in surveys 1,2,4 and 6. Species 2 was confirmed in surveys 2,4 and 6.
- At site, [2,], both species were detected in survey 4, only species 2 was detected in survey 6. No species were confirmed at any survey.
- At site, [3,], only species 2 was detected in surveys 1 and 4. Species 1 was confirmed in survey 1.
- At site, [4,], both species were detected in survey 1, only species 2 was detected in surveys 2,3,4 and 6. Both species 1 were confirmed in survey 1, only species 2 was confirmed in survey 3.

Input to PRESENCE: The additional confirmation data is entered into PRESENCE as the first "survey covariate", and should be named, "conf".

Single-season-Multi-method Model

The multi-method model (Nichols et al. (2008). [9]) extends the single season model by allowing detection probabilities to be different for different methods of observation. This allows the computation of an additional parameter, θ which is the probability that individuals are available for detection at the site, given that they are present.

Parameters:

- ψ - probability that the area is occupied by the species,
- θ^x - probability that individuals are available for detection using method *x*, given presence,
- p^x_i - probability of detecting species using method *x* in survey i,

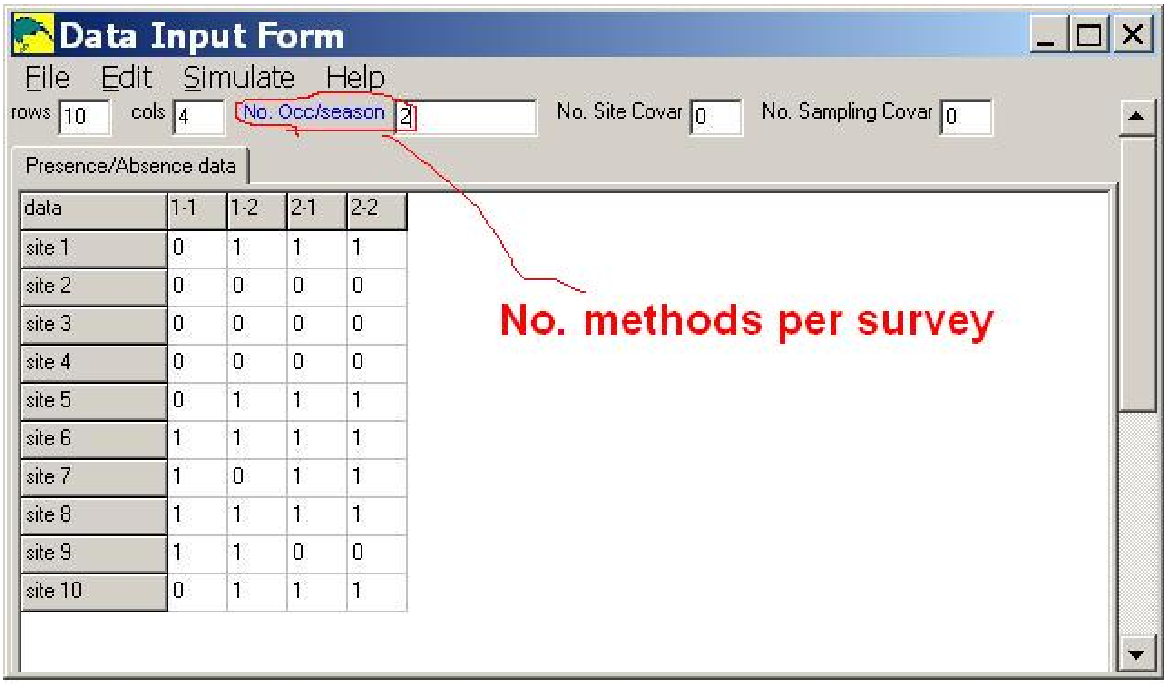
Data input for this model is as follows:

```
site1  h^1,1 h^1,2 h^1,3... h^2,1 h^2,2 h^2,3...
site2  h^1,1 h^1,2 h^1,3... h^2,1 h^2,2 h^2,3...
site3  h^1,1 h^1,2 h^1,3... h^2,1 h^2,2 h^2,3...
:
:
```

where $h^{i,j}$ = 1 if detection at site for survey i, method j; $h^{i,j}$ = 0 if no detection

Additionally, the number of methods per survey is specified in the '#meth/srvy' input box, when entering data.

In the sample input below, there are 10 sites, 2 surveys and 2 methods per survey.



Notice that the column labels (1-1,1-2,2-1,2-2) indicate survey number and method number.

Single-season-Multi-state Model

In the multi-state model (MacKenzie et al., 2009) [5], two kinds of detections are recorded. Detections where only adults observed are recorded as '1' in the data, and detections of known breeding adults (adults seen with young) are recorded as '2' in the data. This allows the computation of an additional parameter, *R* which is the probability that adults breed, given that they are present.

Parameters:

- ψ - probability that the area is occupied by the species,
- *R* - probability that breeders are present, given the area is occupied by the species,
- p^{1}_{i} - probability of detecting adult non-breeders in survey *i*,
- p^{2}_{i} - probability of detecting adult breeders in survey *i*,
- δ_i - probability of correctly identifying breeders (seeing young with adults) in survey i, given presence,

Input data for this model is in the same form as the single-species, single-season model except that breeding status ('1'=adults only, or '2'=adults and young) is recorded instead of presence ('1').

A more general multi-state model allows for more than just two occupied states. In this case, input consists of:

- 0 = no detection
- 1 = detection of species in state 1 (eg., occupied but no breeding detected)
- 2 = detection of species in state 2 (eg., occupied and breeding detected)
- 3 = detection of species in state 3
- : :

Parameters under this parameterization:

- ψ^1 - probability that the area is occupied by the species in state 1,
  - ψ^2 - probability that the area is occupied by the species in state 2,
  - ψ^3 - probability that the area is occupied by the species in state 3,
  - : - :,
  - p^{11}_i - probability of observing species in state 1, given it's true state is 1 in survey *i*,
  - p^{12}_i - probability of ovserving species in state 1, given it's true state is 2 in survey *i*,
  - p^{13}_i - probability of ovserving species in state 1, given it's true state is 3 in survey *i*,
  - : - :
  - p^{21}_i - probability of observing species in state 2, given it's true state is 1 in survey *i*,
  - p^{22}_i - probability of ovserving species in state 2, given it's true state is 2 in survey *i*,
  - p^{23}_i - probability of ovserving species in state 2, given it's true state is 3 in survey *i*,
  - : - :
  - p^{31}_i - probability of observing species in state 3, given it's true state is 1 in survey *i*,
  - p^{32}_i - probability of ovserving species in state 3, given it's true state is 2 in survey *i*,
  - p^{33}_i - probability of ovserving species in state 3, given it's true state is 3 in survey *i*,
  - : - :

  In order for the multi-state model to be identifiable, constraints must be made on the parameters.

  1. higher-order states must imply presence at lower-order states (eg., breeding implies occupancy, but not vice-versa)
  2. Due to the 1^{st} constraint, many detection probabilities need to be fixed to zero (eg., p^{21}=0 -> cannot detect breeding if true state = no breeding (1)).

This model is a special case of the multi-season-multi-state model and can be run in PRESENCE as a multi-season-multi-state model with only one season.

Royle N-Mixture Model

The Royle N-Mixture model (Royle, 2004) [10] estimates population size from temporally replicated point-count data at a number of sample sites. The variation in these point-counts provides information about the distribution of site-specific population size (N). Input data for this model are the counts of the number of individuals observed at each survey (instead of the usual '1' or '0') at each sample site.

Parameters estimated under the assumption of a Poisson distribution:

- λ - population density (per site),
- r - probability of detection (per individual of the species) per survey

---

Multiple Season Model or Dynamic Occupancy model

The multiple season model (MacKenzie et al., 2003) [2] extends the single season model by introducing two additional parameters, $\varepsilon[t]$ and $\gamma[t]$. These parameters are, respectively, the probability a species becomes locally extinct or colonizes a site between seasons t and t+1.

Parameters:

- ψ - probability that the area is occupied by the species,
- p_i - probability of detecting species in survey i, given presence.

- γ_i - probability of unoccupied site being colonized between seasons i and i+1.
- ε_i - probability of occupied site going extinct between seasons i and i+1.

For example, if the detection history 101 000 was observed at a site (denoting the species was detected in the first and third survey of the site in the first season; not detected otherwise), the probability of this occurring could be expressed as;

$\psi \times p_{1,1}(1-p_{1,2})p_{1,3} \times \{(1-\varepsilon_1)\ (1-p_{2,1})\ (1-p_{2,2})\ (1-p_{2,3}) + \varepsilon_1\}$.

This represents the fact that after the first season, the species may have not gone locally extinct $(1-\varepsilon_1)$, but was undetected in the 3 surveys in season 2 $((1-p_{2,1})(1-p_{2,2})(1-p_{2,3}))$ or the species did go locally extinct (ε_1) between the first and second seasons.

The model may also be reparameterized in terms of ψ_t and ε_t; or ψ_t and γ_t, as in some situations this may be a more meaningful parameterization (in terms of overall occupancy) than in terms of the underlying processes. As in the single season model, parameters may be functions of covariates using the logit link.

Note this model does not allow for a so-called "rescue effect", where the local extinction of a colony is "rescued" by the re-colonization of the site before the unoccupied site can be observed (i.e., the site becomes unoccupied then re-occupied all between a single season). Such an effect is sometimes included in metapopulation models, however while a rescue effect is biologically plausible, it can not be estimated (without some potentially unrealistic strict assumptions) from the type of data we are considering here, nor from the type of data often collected in metapopulation studies. The main argument for not including a rescue effect is: why should the rescue of the colony be limited to an arbitrary single event, when possibly there may be a number of opportunities between two seasons for the rescue to occur? To reduce the possibility of having unobserved changes in the occupancy state of sites, the sampling scheme should be designed to reflect the appropriate time scale of the system under study.

Alternate Parameterizations

The initial parameterizaton uses a single initial occupancy paramter, k-1 extinction parameters (assuming k seasons), k-1 colonization parameters, and T detection parameters (assuming T surveys). Once these parameters are estimated, other quantities of interest can be computed. Occupancy in other seasons can be computed as:

$$\psi_2 = \psi_{initial}*(1-\varepsilon_1) + (1-\psi_{initial})*\gamma_1$$
$$\psi_3 = \psi_2*(1-\varepsilon_2) + (1-\psi_2)*\gamma_2$$
$$\psi_4 = \psi_3*(1-\varepsilon_3) + (1-\psi_3)*\gamma_3$$

: : : : :

Sometimes, it is desirable to model seasonal occupancy as a function of some covariates. Since seasonal occupancy is computed from ε_i and γ_i, this cannot be done with these parameters.

An alternate parameterization in PRESENCE uses k occupancy parameters, k-1 extinction parameters, and T detection parameters. The k-1 colonization parameters are then computed from the seasonal ψ's and ε's by solving the above equations for γ_i.

$$\gamma_1 = \frac{\psi_2 - \psi_{initial}*(1-\varepsilon_1)}{(1-\psi_{initial})}$$

By selecting this parameterization, it's now possible to build a model where seasonal occupancy (ψ_i) is a function of a seasonal covariate.

Similarly, we could have estimated the colonization parameters and computed the extinction parameters. This parameterization is sometimes useful if the above parameterization fails to converge on reasonable estimates.

Finally, PRESENCE can model extinction and colonization in such a way that the proportion that go locally extinct is the same as the proportion that don't colonize (ε=1-γ).

Multi-season-staggered-entry model

This model relaxes closure assumption such that a site may locally colonize and go locally extinct once during the surveys (ie., delayed arrival and/or early departure). It is describred by Kendall et. al 2013.[17] and is a simple extension of the single-season-staggered-entry model. Input data is of the same format as the multi-season model.

Parameters:

- ψ^A - probability that the area is occupied by the species,
- b_i - probability of species entering between surveys i and i+1, given not entered yet,
- d_i - probability of species departing between surveys i and i+1, given presence,
- p_i - probability of detecting species in survey i, given presence.
- γ_i - probability of unoccupied site being colonized between seasons i and i+1.
- ε_i - probability of occupied site going extinct between seasons i and i+1.

Multi-season False-positive detection model

The False-positive detection model (Miller et. al. (2013). [17]) extends the single season-false-positive model to allow change in occupancy between seasons. Input format is similar to the single-season-false-positive model.

Parameters:

- ψ - probability that the area is occupied by the species,
- $p11_i$ - probability that the species is detected, given the site is occupied.
- $p10_i$ - probability that the species is detected, given the site is **un**occupied.
- b_i - probability of an 'assured' detection, given the site is occupied and a detection occured.
- γ_i - probability of unoccupied site being colonized between seasons i and i+1.
- ε_i - probability of occupied site going extinct between seasons i and i+1.

Multi-season Spatial Dependence Model

This model is an extension of the Single-season Spatial Dependence model. The parameters for the single-season Spatial Dependence model are replicated for each season, with additional parameters for colonization (γ) and extinction (ε) for each interval between seasons.

Auto-logistic models

With the multi-season model, it may be of interest to model extinction or colonization between seasons *t-1* and *t* as a function of site occupancy for neighboring sites in season *t-1*.

If you're willing to assume that all sites are 'neighbors' of each other (i.e., if the individual organisms can move anywhere within the study area), then this model can be run by simply using *psi1* as a covariate name in the design matrix. When PRESENCE goes to compute colonization or extinction, it will compute the average "conditional" occupancy of neighboring sites in season *t-1* and use that value as a covariate in computing colonization/extinction in season *t*. By specifying "upsi" instead of "psi1", PRESENCE will compute average "unconditional" occupancy of neighboring sites instead of "conditional" occupancy.

Note: "Conditional occupancy" in this case, means conditional on the detection history for the season, not on the entire detection history.

In the case where it is desired to define neighbors specifically for each site, PRESENCE can read a file which defines the neighbors of each site. This file should be a text file containing 1's and 0's where 1 denotes that site *s* is a neighbor of site *r*. The format of the file is rows of contigous 1's and 0's, one row for each focal site. Each row should contain a string of *k* characters (where *k* = number of sites in the study area). For example if there were 10 sites, the neighbor file might look like this:

```
0100000000
1000000000
0001100000
0010100000
0011000000
0000001111
0000010111
0000011011
0000011101
0000011110
```

In this example, sites 1 and 2 are neighbors of each other, sites 3,4,5 are neighbors, and sites 6-10 are neighbors. So, if colonization between seasons *t-1* and *t* is modeled as a function of average neighborhood occupancy at time *t-1*, colonization for site 2 in season *t* will be a function of the average occupancy for sites 1 and 2 in season *t-1*.

In some cases, sites may not all be of the same size or habitat quality. In these cases, it would be preferable to weight the average occupancy of neighboring sites by a value indicating the size/quality of each neighboring site. For example, if all sites except site 5 are nearly the same size, but site 5 is twice as large as the other sites, we would want the occupancy for site 5 to have more influence on colonization/extinction of other sites than the occupancy of other sites. So, the neighbor file would be:

```
0100000000 1.0
1000000000 1.0
0001100000 1.0
0010100000 1.0
0011000000 2.0
0000001111 1.0
0000010111 1.0
0000011011 1.0
0000011101 1.0
0000011110 1.0
```

In this example, sites 3,4,5 are neighbors, but the average occupancy used in the computation of colonization for those sites will be computed as:

$$\mathrm{logit}(\gamma_{site4,t}) = \beta_0 + \beta_1 * X_{site4,t-1} + \ldots$$

where X is the auto-logistic covariate and is computed as:

$$X_{site4,t-1} = (W_{site3} * \psi_{site3,t-1} + W_{site5} * \psi_{site5,t-1}) / (W_{site3} + W_{site5})$$

Using the weights in the example:

$X_{site4,t-1} = (1.0 * \psi_{site3,t-1} + 2.0 * \psi_{site5,t-1}) / (1.0 + 2.0)$

(note: X = auto-logistic covariate and W = weight.)

If PRESENCE finds this auto-logistic covariate name, 'psi1', in the design matrix, it will ask for a neighbor text file. If you don't specifiy a file, PRESENCE will assume all sites are neighbors of each other and all have equal weight (as described initially). If a file is specified, the filename will be saved in the results file (*.pa3), and will be used in all future auto-logistic models. If the file is specified and does not contain a column of weights, all sites will be assumed to have equal weight.

Multi-season-Multi-state Model

This is an extension of the Single-season multi-state model. After the first season, occupancy status can change between each season.

Parameters:

- ψ - probability that the site is occupied initially by the species,
- *R* - probability that breeders are present initially, given the site is occupied by the species,
- $C\psi^0_i$ - probability that the site is occupied in season *i*, conditional that the site was unoccupied (state 0) in season *i-1*,
- $C\psi^1_i$ - probability that the site is occupied in season *i*, conditional that the site was occupied with no breeding (state 1) in season *i-1*,
- $C\psi^2_i$ - probability that the site is occupied in season *i*, conditional that the site was occupied with breeding (state 2) in season *i-1*,
- CR^0_i - probability that breeders are present in season *i*, conditional that the site was unoccupied (state 0) in season *i-1*,
- CR^1_i - probability that breeders are present in season *i*, conditional that the site was occupied with no breeding (state 1) in season *i-1*,
- CR^2_i - probability that breeders are present in season *i*, conditional that the site was occupied with breeders (state 2) in season *i-1*,
- p^1_i - probability of detecting adult non-breeders in survey *i*,
- p^2_i - probability of detecting adult breeders in survey *i*,
- δ_i - probability of correctly identifying breeders (seeing young with adults) in survey i, given presence,

A more general way of parameterizing this model is as follows:

- ψ^{rs}_i - probability that the area is in state *s* in survey *i+1*, given it was in state *r* in survey *i*. (unoccupied=state 0)
- p^{xy}_i - probability of observing species in state *x*, given it's true state is *y* in survey *i*,

Like the single-season-multi-state model, the parameter p^{21}_i would be zero since it would be impossible to observe breeding (state 2) if the species is in state 1 (non-breeding).

So, the first parameterization assumes that there are only 3 occupancy states:

- 0 = unoccupied
- 1 = occupied with no breeding
- 2 = occupied with breeding

The other parameterization can also be used and the relationship between the parameters is:

- p^{10}_i = 0 (impossible to detect species (state 1) if the species is not present (state 0).
- p^{20}_i = 0 (impossible to observe breeding (state 2) if the species is not present (state 0).
- $p^{11}_i = p^{1}_i$ (prob of detecting non-breeders, given only non-breeders present).
- p^{21}_i = 0 (impossible to observe breeding (state 2) if the species is not breeding at site (state 1).
- $p^{12}_i = p^{1}_i * (1-\delta_i)$ (prob of detecting only non-breeders when breeding is taking place)
- $p^{22}_i = p^{1}_i * \delta_i$ (prob of observing breeding when breeding is taking place)
- $\psi^{01}_1 = \psi * (1\text{-}R)$ (initially occupied without breeding)
- $\psi^{02}_1 = \psi * R$ (initially occupied with breeding)
- $\psi^{01}_2 = C\psi^0_2 * (1\text{-}CR^0_2)$ (transition from unoccupied to occupied without breeding from occasion 1 to 2)
- $\psi^{02}_2 = C\psi^0_2 * CR^0_2$ (transition from unoccupied to occupied with breeding from occasion 1 to 2)
- $\psi^{10}_2 = 1\text{-}C\psi^1_2$ (transition from occupied to unoccupied from occasion 1 to 2)
- $\psi^{20}_2 = 1\text{-}C\psi^2_2$ (transition from breeding to unoccupied from occasion 1 to 2)
- $\psi^{21}_2 = C\psi^2_2 * (1\text{-}CR^2_2)$ (transition from breeding to non-breeding from occasion 1 to 2)
- $\psi^{22}_2 = C\psi^2_2 * CR^2_2$ (transition from breeding to breeding from occasion 1 to 2)

This parameterization is more general than the first parameterization, but needs constraints like the ones mentioned above in order to be identifiable. The advantage of this parameterization is that it is possible to have more than 3 occupancy states.

Integrated-habitat-occupancy models

This model generates estimates of changes in occupancy state in relation to changes in habitat state.

Input data consists of the following codes:

0=species not detected at site, habitat state=A,
1=species detected at site, habitat state=A,
2=species not detected at site, habitat state=B,
3=species detected at site, habitat state=B.

Parameters:

- π^X - probability that habitat is initally in state X,
- ψ^X - probability that the site is occupied and habitat is in state X,
- $\varepsilon^{[X,Y]}$ - probability that the site goes locally extinct, given habitat transitions from state X to Y,
- $\gamma^{[X,Y]}$ - probability that the site locally colonizes, given habitat transitions from state X to Y,
- $\eta^{[0,X,Y]}$ - probability that the habitat transitions from X to Y, given site not occupied in previous survey,
- $\eta^{[1,X,Y]}$ - probability that the habitat transitions from X to Y, given occupancy,
- p^X - probability of detecting species, given habitat is in state X.

Note: X = habitat state can be A or B. All parameters except π and ψ can be indexed by survey (subscript i,j in multi-season model framework).

Multi-season Two-Species Model

This is an extension of the two-species model (MacKenzie et al., 2004) [16]allowing the computation of occupancy, colonization, extinction and detection parameters of two species along with conditional probabilities when the other species is present or detected.

Parameters:

- ψ^{A} - probability that the area is initially occupied by species *A*
- ψ^{BA} - probability that the area is initially occupied by species *B*, given that species A is also present
- ψ^{Ba} - probability that the area is initially occupied by species *B*, given that species A is not present
- γ_t^{AB} - probability that the area is colonized by species *A* in the interval, t,t+1, given species B is present in survey t
- γ_t^{Ab} - probability that the area is colonized by species *A* in the interval, t,t+1, given species B is not present in survey t
- γ_t^{BAA} - probability that the area is colonized by species *B* in the interval, t,t+1, given species A is present in survey t and persists in the interval, t,t+1
- γ_t^{BAa} - probability that the area is colonized by species *B* in the interval, t,t+1, given species A is present in survey t and species A goes extinct in the interval, t,t+1
- γ_t^{BaA} - probability that the area is colonized by species *B* in the interval, t,t+1, given species A is not present in survey t and species A colonizes in the interval, t,t+1
- γ_t^{Baa} - probability that the area is colonized by species *B* in the interval, t,t+1, given species A is not present in survey t and species A does not colonize in the interval, t,t+1
- ε_t^{AB} - probability that the area goes extinct by species *A* in the interval, t,t+1, given species B is present in survey t
- ε_t^{Ab} - probability that the area goes extinct by species *A* in the interval, t,t+1, given species B is not present in survey t
- ε_t^{BAA} - probability that the area goes extinct by species *B* in the interval, t,t+1, given species A is present in survey t and species A persists in the interval, t,t+1
- ε_t^{BAa} - probability that the area goes extinct by species *B* in the interval, t,t+1, given species A is present in survey t and species A goes extinct in the interval, t,t+1
- ε_t^{Baa} - probability that the area goes extinct by species *B* in the interval, t,t+1, given species A is not present in survey t and species A does not colonize in the interval, t,t+1
- p^{A} - probability of detecting species *A*, given species *B* is not present
- p^{B} - probability of detecting species *B*, given species *A* is not present
- r^{A} - probability of detecting species *A*, given both are present
- r^{BA} - probability of detecting species *B*, given both are present and species A detected
- r^{Ba} - probability of detecting species *B*, given both are present and species A not detected

Input data for this model is in the same form as the single-species, single-season model except that the first half of the detection history records are assumed to be species A, and the second half of the records are assumed to be species B. So, if there are 60 sites, the input would consist of 120 detection history records. Records 1-60 would be the site-detection history records for sites 1-60, species A, and records 61-120 would be the site-detection history records for sites 1-60, species B.

Alternate Input for 2-species model-
Instead of repeating each site for each species, the following codes can be used for this model:

- 0 - neither species detected
- 1 - only species A detected
- 2 - only species B detected

- 3 - both species detected

If there are any 2's or 3's in the data, PRESENCE will assume this form of input.

WinBugs Analysis

A rudimentry capability has been programmed into PRESENCE to do an analysis for select models using the Baysian statistical package, WinBugs. For certain model types, a button will appear on the 'Run model' dialog window with the caption 'Run WinBugs'. When this button is clicked, PRESENCE will generate the model code and data files needed by WinBugs to analyze the data. PRESENCE will then call WinBugs to do the analysis, then display a Notepad window with the results. Eventually, PRESENCE will be able to do more models, but at the moment, only the single-season model is available. Note: WinBugs must be installed for this feature to work.

Model Run Options

- List Input Data - Prints input data in output file
- Supply initial values - Displays a window where initial values of beta parameters can be input
- Set digits in estimates - Prompts for number of digits past the decimal point that the likelihood value doesn't change in the optimization routine before optimization stops
- Set function evaluations - Prompts for the maximum number of function calls before aborting optimization
- Print V-C matrix - Prints the variance-covariance matrix of real parameters in the output file
- Don't compute V-C matrix - Speeds up run-time by not computing the variance-covariance matrix of beta parameters
- Bootstrap V-C matrix - Prompts for whether the variance-covariance matrix should be computed via bootstrap (and asks for the number of bootstrap samples) instead of using the Hessian
- Assess Model Fit - Prompts whether a goodness-of-fit (gof) test should be performed
  - Note: gof test cannot be performed on all models
  - Note: data cannot be summarized for gof test
- Use simplex algorithm for starting values - Prompts for whether to use simplex algorithm to get better starting values
- Set max estimates to print - Prompts for the maximum number of individual site/survey estimates of real parameters to print in the output file
- Use Complimentry log-log link for Psi - Compute real estimate of occupancy using log-log link, instead of logistic link

---

Single Season Output

The results for fitting the single season model to the data are stored in the results database. To view the output of a specific model, position the cursor over the desired model name, click with the right mouse button, then select 'view output' with the left mouse button. If the 'list data' option had been selected, the input data will appear at the beginning of the output. Next, the number of sites, sampling occasions and missing observations in the dataset; followed by the number of parameters in the model, twice the negative log-likelihood and Akaike's Information Criterion (AIC), (e.g., For **simple single-season models**, after the model AIC has been output, the estimated coefficients for the logistic model and their variance-covariance matrix are printed. These are the 'beta' or untransformed estimates which can be used to compute the 'real' parameters (psi,p) in the model.

```
Untransformed Estimates of coefficients for covariates (Beta's)
===========================================================================
                                          estimate   std.error
A1     : psi                              1.098612 (0.248070)
B1     : p1                              -0.000000 (0.112774)
```

```
Variance-Covariance Matrix of Untransformed estimates (Beta's):
              A1         B1
     A1    0.061539  -0.004103
     B1   -0.004103   0.012718
------------------------------

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

   Individual Site estimates of <Psi>
        Site         Survey         Psi     Std.err     95% conf. interval
1             0         survey_1:  0.7500   0.0465     0.6485 - 0.8299

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

   Individual Site estimates of <p>
        Site         Survey           p     Std.err     95% conf. interval
1             0         survey_1:  0.5000   0.0282     0.4450 - 0.5550
1             0         survey_2:  0.5000   0.0282     0.4450 - 0.5550
1             0         survey_3:  0.5000   0.0282     0.4450 - 0.5550
1             0         survey_4:  0.5000   0.0282     0.4450 - 0.5550
1             0         survey_5:  0.5000   0.0282     0.4450 - 0.5550

============================================================
 DERIVED parameter - Psi-conditional : [Pr(occ | detection history)]

        Site                    psi-cond  Std.err     95% conf. interval
     1        0                   0.0857  0.0315     0.0240 - 0.1474
     2   site_2                   1.0000  0.0000     1.0000 - 1.0000
     3   site_3                   1.0000  0.0000     1.0000 - 1.0000
     4   site_4                   1.0000  0.0000     1.0000 - 1.0000
```

Note that if no covariates are used in the estimation of a real parameter, all sites will have the same value for that parameter and PRESENCE will only print the estimate for the 1st site.

After printing the 'real' parameters, a 'derived' parameter (psi-conditional) is printed. This parameter is the probability that a site is occupied, given it's particular detection history. So, all sites where a detection occured, will have a value of 1.0 for this parameter. Sites with no detections will have a value of less than or equal the unconditional occupancy estimate (psi). "Conditional occupancy" can be computed as:

$$\psi^c = \frac{\psi\ \Pi(1-p_j)}{\psi\ \Pi(1-p_j) + 1-\psi} \quad \text{if no detections, = 1 if at least 1 detection}$$

This parameter is useful for generating maps of species occurence.

Multiple Season Output

The results for fitting the multiple season model to the data will appear in a Notepad window. As for single season models, the output begins with a listing of the input data (if desired), followed by the number of sites, sampling occasions and missing observations in the dataset. The design matrices are printed followed by the number of parameters in the model, twice the negative log-likelihood and Akaike's Information Criterion (AIC). The untransformed parameter estimates and their associated variance-covariance matrix are printed, followed by the 'real' parameter estimates.

```
Untransformed Estimates of coefficients for covariates (Beta's)
==============================================================================
                                            estimate   std.error
A1      : psi                                1.098612 (0.411930)
B1      : gam1                              -2.197225 (2.394989)
C1      : eps1                              -1.386294 (0.468723)
D1      : p1                                -0.000000 (0.159541)

Variance-Covariance Matrix of Untransformed estimates (Beta's):
```

```
            A1         B1         C1         D1
     A1    0.169686  -0.670645  -0.036365  -0.033936
     B1   -0.670645   5.735971   0.459619   0.191898
     C1   -0.036365   0.459619   0.219701   0.027272
     D1   -0.033936   0.191898   0.027272   0.025453
------------------------------
Model has been fit using the logistic link.

   Individual Site estimates of Psi:
        Site          Survey          Psi     Std.err     95% conf. interval
1            0           survey_1:  0.7500   0.0772     0.5723 - 0.8706

   Individual Site estimates of Gamma:
        Site          Survey        Gamma     Std.err     95% conf. interval
1            0           survey_1:  0.1000   0.2155     0.0010 - 0.9239

   Individual Site estimates of Eps:
        Site          Survey          Eps     Std.err     95% conf. interval
1            0           survey_1:  0.2000   0.0750     0.0907 - 0.3852

   Individual Site estimates of p:
        Site          Survey            p     Std.err     95% conf. interval
1            0           survey_1:  0.5000   0.0399     0.4225 - 0.5775
1            0           survey_2:  0.5000   0.0399     0.4225 - 0.5775
1            0           survey_3:  0.5000   0.0399     0.4225 - 0.5775
1            0           survey_4:  0.5000   0.0399     0.4225 - 0.5775
1            0           survey_5:  0.5000   0.0399     0.4225 - 0.5775
```

Tools and Settings

- Simulation tool - Instructional tool to give users a general feel for how the model of MacKenzie et al. (2002)[1] performs under a specific set of circumstances and sampling designs.
- Check website for new version - Launches a browser window with the PRESENCE home page so you can check if your version is up-to=date.
- Plot psi vs covariates - Generates a graph of psi estimates on the y-axis versus covariates on the x-axis for a user-specified logit equation.
- Change web-browser - Allows PRESENCE to use a different web-browser for display of help items and web-pages. When dialog box appears, locate alternate browser (eg., "c:\Program Files\Mozilla Firefox\firefox.exe").
- Change text-editor - Allows PRESENCE to use a different text-editor for display of model output. When dialog box appears, locate alternate editor (eg., "c:\Program Files (x86)\Notepad++\notepad++.exe").
- Specify location of OpenBugs - Allows PRESENCE to find OpenBugs.exe for running simple Baysian models. When dialog box appears, locate the OpenBugs.exe file (eg., "c:\Program Files (x86)\OpenBugs\OpenBugs.exe").
- Change c-hat - Once Goodness-of-fit is assessed, enter the computed c-hat value to modify results table to use QAIC instead of AIC.
- Change effective sample size - When a value is entered for effective sample size (ESS), results table will display AICc instead of AIC (or QAICc instead of QAIC).
- Add 1 to #parms when c-hat>1 - increments parameter count in each model to account for model uncertainty.
- Model averaging - computes model-averaged estimates of site-specific occupancy or detection.
- Set #estimates to print by default - By default, only 200 site estimates are printed in the output file (to save paper). To see all estimates, enter a larger number here.
- Rebuild project from output files - Scans files in current folder for PRESENCE model output files and creates the model summary table.

Single Season Simulation

This simple simulation routine is included so that users may get a general feel for how the model of MacKenzie et al. (2002)[1] performs under a specific set of circumstances and sampling designs. Scenarios may either be entered from a tab-delimited ASCII text file (see below for details), or by entering the scenario directly.

Where the previously described simulation procedure is intended as a basic learning tool, this procedure is designed to address a specific question.

There are two general sampling designs that can be investigated; sampling only a subset of sites more intensively to estimate detection probabilities; or halting the repeated sampling of sites after the species is first detected. Both designs are compatible with the MacKenzie et al. (2002)[1] model. A single-group model with constant p is fit to the simulated data. Results are written to a file named 'presence.out' and loaded into the *Notepad* editor when completed.

This simulation file should be set up as follows (see SimExample.txt).

The first line should consist of 4 integer values;

- the total number of sites(N)
- the number of sites to be sampled more intensively (NI)
- the number of surveys conducted at the subset of sites (T)
- the number of surveys conducted at the other sites (T0)

The next N lines of the file hold the true occupancy and detection probabilities for each site. The first column in each line is the occupancy probability, and the following T columns contain the probability of detecting the species (given presence) during each survey.

The first NI of these lines represent the sites that will be sampled more intensively. For the remaining sites, if T0 < T then PRESENCE is still expecting to read in T detection probabilities, however these will not be used.

The final line of the file consists of 3 integer values;

- the number of unknown groups (see single season analysis)
- whether surveying will halt after first detection (0=no, 1=yes)
- the number of simulations

SimExample.txt - sample simulation input file

```
20      10      5       3
0.8     0.2     0.4     0.3     0.2     0.6
0.8     0.2     0.4     0.3     0.2     0.6
0.8     0.2     0.4     0.3     0.2     0.6
0.8     0.2     0.4     0.3     0.2     0.6
0.8     0.2     0.4     0.3     0.2     0.6
0.8     0.2     0.4     0.3     0.2     0.6
0.8     0.2     0.4     0.3     0.2     0.6
0.8     0.2     0.4     0.3     0.2     0.6
0.8     0.2     0.4     0.3     0.2     0.6
0.8     0.2     0.4     0.3     0.2     0.6
0.8     0.2     0.4     0.3     0.2     0.6
0.8     0.2     0.4     0.3     0.2     0.6
0.8     0.2     0.4     0.3     0.2     0.6
0.8     0.2     0.4     0.3     0.2     0.6
0.8     0.2     0.4     0.3     0.2     0.6
0.8     0.2     0.4     0.3     0.2     0.6
0.8     0.2     0.4     0.3     0.2     0.6
0.8     0.2     0.4     0.3     0.2     0.6
0.8     0.2     0.4     0.3     0.2     0.6
0.8     0.2     0.4     0.3     0.2     0.6
1       0       500
```

Model averaging

Just as there is uncertainty in model parameters, there is also uncertainty in model selection. There may be several models with (relatively) equal support, according to AIC values. In this case, parameters may be averaged among all models, such that models with greater support have more weight in the computed average than models with little support.

Model averaged parameters may be computed in PRESENCE by selecting the "Model averaging" menu from the "Tools" menu. PRESENCE will obtain estimates for each parameter for each site (and survey if applicable) along with the weight of each model and compute a weighted average estimate. The two options for output are:

- print only model averaged estimate and standard error for each site/survey
- print all estimates in each model used in the computation of the model-averaged estimate

---

Help menu

- Online book - Opens web-browser and loads Donovan and Hines online spreadsheet book
- Single-season Tutorial1 - Opens web-browser and loads step-by-step instructions for analyzing a sample data file with single-season models.
- Single-season Tutorial2 - Opens web-browser and loads step-by-step instructions for analyzing a sample data file with single-season models.
- Multi-season Tutorial1 - Opens web-browser and loads step-by-step instructions for analyzing a sample data file with multi-season models.
- Multi-season Tutorial2 - Opens web-browser and loads step-by-step instructions for analyzing a sample data file with multi-season models.
- Check web for upcoming workshops - Opens web-browser and loads a web-page with a list of upcoming occupancy modeling workshops.
- Occupancy Users forum (Phidot.org) - Opens web-browser and loads PHIDOT.ORG user-group forum.
- Attitude adjustment - Displays a graphic intended to relax users who have run one too-many models :-)

Problems/Questions

The most frequent question/problem that occurs is about a message in the output warning that the optimization routine has possibly not reached the maximum likelihood value. The program then prints the number of significant digits achived when the optimization stopped. This situation occurs when the optimization routine is attempting to find the maximum of a function, when the function is relatively 'flat' (thinking in two-dimensions). If you look at the logit transformation described in this help file, you'll see that this can occur if a parameter is near zero or one. When a parameter is near zero, the 'untransformed' or 'beta' parameter associated with it is a large negative number. To get the 'real' parameter, the untransformed parameter is plugged into the logit equation (exp(x)/(1+exp(x))). So, plugging in -30 for x in this equation gives almost the same value as -40. This causes the optimization routine to think the function is 'flat' and gives the warning message.

Examining many simulated and real datasets, we have found that this message can be safely ignored if the number of significant digits is 3 or larger. The number of significant digits it reports is not the number of digits you can trust in the parmeter estimates. We have found that even when it reports 2 significant digits, the esitmates are accurate to 4 or more decimal places.

When the message occurs with the number of significant digits less than two, it usually indicates insufficient data for the desired model, or model overparameterization. This is sometimes accompanied by a warning about the variance-covariance matrix. If this happens, the model may need to be simplified.

In some cases, poor starting values for the parameters can cause the problems noted above. This can be solved by giving better initial values to the program when running the model. For example, if detection probabilities are very small, and the default starting values of 0.5 are far away from the final expected parmaeter values, the optimization routine may fail. The solution would be to input small initial values (on the logit scale) for the model so the optimization routine does not have to search very far. Since simpler models converge more readily than complex ones, it is usually best to start with simple models, so you have starting values for complex ones if needed.

Another indication of a problem with convergence is strange standard error values. PRESENCE may or may not print the error message described above, but the model may still be over-parameterized. It is very important to check estimates and standard errors for all models to make sure they are reasonable. Note that large standard errors of the 'real' parameters (psi,p,eps,gam,...) indicate problems, but large standard errors for the untransformed (beta) parameters are not necessarily a problem. Particularly, when a beta value is large (>10 or <-10), the standard error is usually large also.

Resources

- A book[8] discussing the latest methods in analyzing presence/absence data surveys has been written and is available at: Elsevier.com

  MacKenzie, D.I., J.D. Nichols, J.A. Royle, K.H. Pollock, L.L. Bailey, and J.E. Hines, 2017. Occupancy Estimation and Modeling: Inferring Patterns and dynamics of Species Occurance, 2nd Edition, Elsevier Publishing, Inc.

- Workshops are conducted several times per year which teach the basic principals of Occupancy Modeling. A list of upcoming workshops is available at Proteus.com

- Another resource for self-help in Occupancy modeling is an on-line publication which explains how occupancy models work via spreadsheets:

  Donovan, T. M. and J. Hines. 2007. Exercises in occupancy modeling and estimation.

- An online forum dedicated to occupancy-related questions has been set up. The general intent is to create a forum where people can exchange ideas and pose questions about species occurrence. Posts about both how to do the modeling in PRESENCE(or other software) and study design issues will be accepted. However,it is expected that before posting a question you will have checked help files, articles etc. for an answer; else risk a public flogging;-)

  To join phidot, go to PHIDOT.ORG and select 'regsiter' near the top of the screen. Once joined, you can set individual preferences, etc.

- A comprehensive introductory on-line manual for program MARK has been written which gives many details of capture-recapture, occupancy modeling and other types of analysis. Methods for analyzing data in PRESENCE are very similar to methods in MARK. In fact, MARK will do many of the same models as PRESENCE. The on-line manual can be found at: http://www.phidot.org/software/mark/docs/book/

- Program GENPRES can be used to simulate (or generate expected value) data under most PRESENCE models. It can be used in the design stage of an occupancy experiment to get an idea of how small/large the variances of estimates might be, given a set of parameters. It is also useful for sensitivity analysis or as an educational tool for learning about the models.

- PRESENCE can be called from the R software package (www.r-project.org), making it possible to write r-scripts to run models on many data-sets or run many models on one data-set. Some R functions are included in the PRESENCE/R-tools folder to perform tasks necessary convert data and run models. Sample code is also included.

Credits/Acknowledgments

PRESENCE was developed by Darryl MacKenzie of Proteus Research & Consulting Ltd. under contract to U.S. Geological Survey as part of their Amphibian Research and Monitoring Initiative.

Versions 2 and up of PRESENCE were developed by Jim Hines of the U.S. Geological Survey.

Currently, We don't know of any bugs in PRESENCE, although that doesn't mean there aren't any (yes, detection probability is less than 1.0!). If you find some, feel free to let us know.

Jim Hines jhines@usgs.gov

Darryl MacKenzie darryl@proteus.co.nz

Appendix

Covariates

Program PRESENCE makes the distinction between two types of covariates. Site-specific covariates are covariates that are constant for a site within a season. Examples would be habitat type, patch size, distance to nearest patch, or generalized weather patterns such as drought or El Niño years. Sampling-occasion covariates are covariates that may change with each survey of a site, for example local environmental conditions such as temperature, precipitation or cloud cover; time of day; or observer. Covariates are entered into the models using the logistic model.

Detection probabilities may be functions of either site-specific or sampling occasion covariates, while all other parameters may be functions of site-specific covariates only.

Sampling-occasion covariates may be missing, and are assumed to correspond to a missing detection/nondetection observation. When a covariate is being used that has missing values that do not correspond with a missing detection/nondetection observation, the detection/nondetection data is also treated as missing. Site-specific covariates can not have missing values, unless the site was never surveyed during that season.

An important note about continuous covariates! Because of the way the logit-link works, if the average value of a covariate is a long way from zero, then PRESENCE may not be able to find the true maximum likelihood estimates of the model parameters, which will give you bogus results. An indication that there might be a problem is that the estimates themselves look suspicious, the variance-covariance matrix might include a huge value, and/or you get a warning about a non-invertible variance-covariance matrix. The best approach is to transform your data onto another scale which is still meaningful to you. You could divide the covariate values by some constant (i.e., rather than entering 80% humidity as 80.0, use 0.80); subtract the average of the covariates from each observed value (i.e., X* = X - average(X's)); or some combination of the two. Such transformations can be carried out by PRESENCE (in the edit menu) or done easily with a spreadsheet and the modified values pasted back into the Data Window.

Logistic Model or Logit Link

The logistic model can be used to investigate potential relationships between probabilities (the response) and covariates (the explanatory variables), as it ensures response values stay between 0 and 1. The logistic model is defined as;

$\log_e(y/(1-y)) = X\beta$,

where y is the probability; X is a row vector containing the covariate values; and β is a column vector of coefficient values that are to be estimated. An alternative definition for the model is, $y = \exp(X\beta) / (1+\exp(X\beta))$.

Large positive values for Xβ make y tend to 1, while large negative values make y tend to 0. If $X\beta = 0$, then $y = 0.5$.

Literature Cited

[1]MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle and C. A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83(8): 2248-2255.

[2]MacKenzie, D. I., J. D. Nichols, J. E. Hines, M. G. Knutson and A. B. Franklin. Estimating site occupancy, colonization and local extinction probabilities when a species is not detected with certainty. Ecology, 84, 2200-2207.

[3]Pledger, S. 2000. Unified maximum likelihood estimates for closed capture-recapture models using mixtures. Biometrics 56: 434-442.

[4]Burnham, K. P. and D. R. Anderson. 1998. Model selection and inference. Springer-Verlag, New York, USA

[5]MacKenzie, D.I., J.D. Nichols, M.E. Seamans and R.J. Gutiérrez. 2009. Modeling species occurence dynamics with multiple states and imperfect detection, Ecology 90, pp. 823-835.

[6]Royle, J.A. 2004. N-Mixture Models for Estimating Population Size from Spatially Replicated Counts. Biometrics 60, 108-115.

[7]Royle, J.A., and J.D. Nichols. 2003. Estimating Abundance from Repeated Presence-Absence Data or Point Counts. Ecology 84(3):777-790.

[8]MacKenzie, D. I., J. D. Nichols, J. A. Royle, J.A., K. Pollock, L. Bailey and J. E. Hines. 2006. Occupancy Estimation and Modeling - Inferring Patterns and Dynamics of Species Occurrence. Elsevier Publishing.

[9] Nichols, J.D., L.L. Bailey, A.F. O'Connell Jr., N.W.Talancy, E.H.C. Grant, A.T. Gilbert, E.M. Annand, T.P. Husband and J.E. Hines. 2008. Multi-scale occupancy estimation and modelling using multiple detection methods. Journal of Applied Ecology 45(5):1321-1329

[10] Royle, J. A. 2004. N-mixture models for estimating population size from spatially replicated counts. Biometrics 60:108-115.

[11] J. E. Hines, J. D. Nichols, J. A. Royle, D. I. MacKenzie, A. M. Gopalaswamy, N. Samba Kumar, and K. U. Karanth. 2010. Tigers on trails: occupancy modeling for cluster sampling. Ecological Applications 20:1456-1466.

[12] Miller, D.A., J.D. Nichols, B.T. Mcclintock, E.H.C. Grant, L.L. Bailey, and L. Weir. 2011. Improving Occupancy Estimation When Two Types Of Observational Error Occur: Non-Detection And Species Misidentification. Ecology 92:1422-1428.

[13] Kendall, W.L., J.E. Hines, J.D. Nichols, and E.H. Grant. 2013. Relaxing The Closure Assumption In Single-Season Occupancy Models: Staggered Arrival And Departure Times. Ecology 94(3):610-617.

[14] D.I. MacKenzie, J.D. Nichols, L.L. Bailey, and J.E. Hines 2011. An Integrated Model of Habitat and Species Occurrence Dynamics. Methods in Ecology and Evolution 2:612-622.

[15] Royle, J. A., and J. D. Nichols. 2003. Estimating abundance from repeated presence-absence data or point counts. Ecology 84:777-790.

[16]MacKenzie, D.I., L. L. Bailey, and J.D. Nichols. 2004. Investigating species co-occurence patterns when species are detected imperfectly, Journal of Animal Ecology 73, pp. 546-555.

[17] Miller, D.A.W., J.D. Nichols, J.A. Gude, L.N. Rich, K.M. Podruzny, J.E. Hines, and M.S. Mitchell. 2013. Determining occurrence dynamics when false positives occur: estimating the range dynamics of wolves from public survey data. PLoS ONE 8(6): e65808. doi:10.1371/journal.pone.0065808.

[18] Chambert, T., E.H. Campbell-Grant, D.A.W. Miller, J.D. Nichols, K.P. Mulder and A.B. Brand. 2018. *Two-species occupancy modelling accounting for species misidentification and non-detection.* Methods in Ecology and Evolution, https://doi.org/10.1111/2041-210X.12985