

OCCUPANCY ESTIMATION AND MODELING

Overview of occupancy sampling and modeling

The general problem of occupancy (“presence-absence”) sampling involves n sites, over which we take k detection samples. In the simplest case (which we will start with) there is a single season, and the detections are purely detected (1) or not (0) at each sample, where detection means that we have detected 1 or more animals, without regard to how many have been detected (obviously if $N > 1$ the site is still occupied!). The data for an occupancy study over 5 sites might look like this

10101

01000

00000

10000

11111

Note that 1) we can conclude (for a ‘closed’, single-season study) that sites 1, 2, 4, and 5 are occupied. But we cannot conclude that site 3 is not occupied—it may not be, or it may be but there were simply no detections.

Steps for data entry and model running in PRESENCE

This is a quick overview of how to run PRESENCE; for details please refer the [PRESENCE user’s guide](#). I assume that you’ve already installed PRESENCE on your operating system (refer to the user’s guide for details on this).

Running the program

I assume here that you are operating on a Windows machine. To run the program simply click on the PRESENCE icon on the desktop, or go to the taskbar and find the PRESENCE folder, click on the green bird (a Kiwi) in the yellow box. This opens up a graphical user interface (GUI) that will be your main method of inputting data, specifying models, and sorting through output.

Entering data

There are a number of options for entering data (see the user's guide for details). For example, it is possible to enter data directly into the data input form available on the GUI; this is usually only appropriate for very small studies (say <30 sites). In most cases you will want to store and edit your data in a database or spreadsheet, and copy and paste the data (encounters, covariates, etc.) as needed into PRESENCE. That is the approach we will take here.

For basic (single-state, no count data) occupancy problems, the essential data are the *encounters* at each sampling occasion for each site. The data will form a matrix where the rows are sites and the columns are sampling occasions. For each site, a "1" is entered in an occasion column if 1 or more animals are detected, and "0" otherwise. Sometimes no sampling occurs at a site on a particular occasion, in which case "." is used to indicate a missing value (not the same as a non-detection). Thus for 3 sites over 3 occasions we might have

101

000

111

which would indicate detection on the first site at the first and 3rd occasions, no detections on the second site, and detections all 3 occasions on the 3rd site. By contrast

101

0.0

111

would indicate that the second site was not sampled on occasion 2 (so we don't know if a detection would have occurred).

To start a new project, open the GUI and go the 'file' on the main menu and select 'New Project'. This will be where we enter data and specify the essential parameters of the data structure (how many sites, how many occasions, whether there are covariates, etc.) At this point we could open an existing data file in PRESENCE; instead we will copy data from our Excel spreadsheet.

FIRST EXAMPLE – Blue Grosbeaks in Georgia

We applied occupancy analysis to observations of blue grosbeaks (*Guiraca caerulea*) on 41 old fields planted to longleaf pines (*Pinus palustris*) in southern Georgia, USA. Surveys were 500 m transects across each field and were completed three times during the breeding season in 2001. The spreadsheet of the data can be found in the BLGR AEC 501 spreadsheet; for now we are going to ignore the covariates and other information in the spreadsheet, and focus on the encounter histories for each site (columns B, C, and D starting in row 2 and running to row 42 (41 sites total). First, go to the data input form and specify 41 rows, 3 columns, and 3 occasions per season (this will specify that you have 41 sites and 3 occasions, but a single season. Return to the spreadsheet and select the cells in columns B,C, and D, rows 2 through 42 in the spreadsheet and copy them. Then place your cursor in the first cell (site 1 occasion 1), go to 'edit' in the input form, and select "paste" and "paste values". You should wind up with 41 rows of data.

After the data are pasted as above, you must click 'File/Save' from the menu, then click 'File/Close'. ***This is a critical step: PRESENCE will not be able to use data in the form unless it has been saved.*** You will be asked if the last column is a frequency: in this example, answer "no" (each line represents a single site). You will need to provide a title; use "BLGR". You will also need to supply a unique file name for the data. Use blgr.pao. Then select "File /Close".

Return to the dialog screen and click the box "Click to select file" and browse to find the data file you just entered. The data along with design parameters (number of sites, number of occasions/ season, etc.) should now be filled in. Finally, click "Ok" (green check box). You are ready for the next step: running your first occupancy model.

We will start with a single model, specifying constant detection in a single season. Go to the main menu and select "Run", then "Analysis single season". For now ignore the "Fix Parameters" box and the "Options" check bubbles, and concentrate on the list mode models. We are going to run a single model: 1 group, constant P.

Run this model, click 'yes' to append the result to the Results Browser. The results file (blgr.pa3) will be created when you exit PRESENCE. Now both the data and the model results have been saved and are available in this project file. Highlight the model in the results browser and click on the detailed results box (second left) to view the model output. We will focus for now on the results

Proportion of sites occupied (Psi) = 0.8848 (0.0766)

and

detection probabilities (p):

grp	srvy	p	se(p)
-----	------	---	-------

--- ---- -

1 1 0.551317 (0.059854)

Also notice the “naïve occupancy estimate” of 0.8049. This is the estimate we would get by simply looking at the proportion of the 41 sites in which at least 1 detection occurred. Because detection rates are considerably lower than 1, this underestimates true occupancy (some of the “zeros” sites probably were occupied). The estimate of 0.8848 (SE=0.0766) gives a more accurate picture of actual occupancy.

Generalization to allow for sample-specific detection

So far we have considered the simple case, where the probability of occupancy (ψ or ψ_i) is the same among sites, and where detection probabilities (p) are the same for all sites and over all samples. We will relax this assumption a bit by allowing detection probabilities to vary between samples (but within a sample the same among sites); occupancy probabilities remain homogeneous. The result is that we have a more complex model (more parameters), because we will now have a detection probability for each site.

We can illustrate this again with the Blue Grosbeak example, by returning to the data project we already created. If you have not closed this project, we can simply return to the menu and run another model (since the data haven't changed). If you have closed the project, no problem- just open PRESENCE, go to 'File/Open project' and browse to the folder containing the project (it will be named something like blgr_project) and open the project file (blgr.pa3). We now want to run an additional model. We are still in 'Analysis:single-season'; we now want to run “1 group, Survey-specific p”. Run this model and add it to the Results Browser. First, note that the AIC value for this model is higher (by 3.11 AIC units) than that for the simpler model (constant p). This indicates that the simpler model should be preferred based on the “Principle of Parsimony”, since using too many parameters unnecessarily will add to the imprecision of the model estimates. We can still examine the output by opening the detailed output dialog box, where we will observe that the estimate of ψ_i is very similar to the previous model (0.8827 vs. 0.8848) but now we have 3 estimates of p (0.497, 0.607, 0.553) instead of a constant $p=0.551$. So, there is some indication of sample-to-sample variation, but not enough to support the more complex model.

Common Mistakes and Errors

These are common mistakes or errors that arise when running an analysis and indicate some kind of problem that should be checked more thoroughly.

1. Estimate of p (detection) and/or ψ (occupancy) are 0 or 1 (or very close to 0 or 1).
 - a. check #2 below, if it is okay then estimate of p or ψ should be okay
 - b. double check to see if you have no 1's in your data set or all 1's (every site has a detection or no site has a detection)
2. Estimate of the SE (standard error) is very large (e.g., >100) or very small (~ 0)
 - a. double check your data to make sure there are no mistakes
 - b. if no mistakes in data, problem may lie in not enough data to estimate parameters (e.g., only had 5 total sites and very few detections)
3. Var-Cov matrix not identifiable/estimable or is singular
 - a. go through steps for #2 above
4. Missing data
 - a. missing data is handled by a single dot (.) in the data sheet
5. Error in data sheet (incorrect number of visits/surveys, sites, number that is not 0 or 1)...
 - a. double check your data and especially double check that the number of sites and surveys in program presence is accurate