EXERCISE 14: SINGLE-SEASON, SPECIES-INTERACTIONS OCCUPANCY MODELS

In collaboration with Rebecca J. Pfeiffer and Jeremy M. Clark

University of Vermont, Rubenstein School of Environment and Natural Resources

Please cite this work as: Donovan, T. M. and J. Hines. 2007. Exercises in occupancy modeling and estimation. <http://www.uvm.edu/envnr/vtcfwru/spreadsheets/occupancy.htm>

TABLE OF CONTENTS

SINGLE-SEASON, SPECIES-INTERACTIONS OCCUPANCY SPREADSHEET EXERCISE ... 3
- **OBJECTIVES:** ... 3
- **BACKGROUND READING** ... 3
- **MODEL BACKGROUND** ... 3
- **ENCOUNTER HISTORY BASICS** ... 4
- **ENCOUNTER HISTORY PROBABILITIES AND CO-OCCURRENCE PARAMETERS** ... 6
- **OCCUPANCY PARAMETERS** ... 7
- **DETECTION PARAMETERS** ... 11
- **CREATING ENCOUNTER HISTORY PROBABILITIES WITH OCCUPANCY AND DETECTION PARAMETERS** ... 14
- **THE CO-OCCURRENCE LOG LIKELIHOOD** ... 17
- **MODEL OUTPUT** ... 18
- **THE MODEL CHI-SQUARE** ... 19
- **THE SATURATED MODEL** ... 20
- **RUNNING MODELS** ... 20
- **MODEL 1. MODEL PSIA, PSIB, PSIAB, PA (T) PB (T) PAB(T)** ... 21
- **MODEL 2. MODEL PSIA, PSIB, PSIAB, DETECTION (.)** ... 25
- **MODEL 3. MODEL PSI A, PSI B, DETECTION(.)** ... 27
- **RUNNING OTHER MODELS** ... 29
- **SIMULATING CO-OCCURRENCE DATA** ... 31
- **CREATING DATA FROM EXPECTED VALUES** ... 32
- **CREATING DATA WITH STOCHASTICITY** ... 33
- **PRESENCE INPUT FILES** ... 38

SINGLE-SEASON, SPECIES-INTERACTIONS OCCUPANCY MODELS IN PRESENCE ... 39
- **GETTING STARTED** ... 39
- **MODEL 1. MODEL PSIA, PSIB, PSIAB, PA (T) PB (T) PAB(T)** ... 41
- **MODEL 2. MODEL PSIA, PSIB, PSIAB, DETECTION** ... 42
- **MODEL 3. MODEL PSI A, PSI B, DETECTION(.)** ... 52

SINGLE-SEASON, SPECIES-INTERACTIONS OCCUPANCY SPREADSHEET EXERCISE

OBJECTIVES:

- To learn and understand the co-occurrence occupancy model, and how it fits into a multinomial maximum likelihood analysis.
- To use Solver to find the maximum likelihood estimates for the probability of detection of two species, the probability of site occupancy of two species, and how the species influence each other's detection and occupancy probabilities.
- To assess model fit by comparing the $-2\text{Log}_e\text{L}$ of the model to the $-2\text{Log}_e\text{L}$ of the saturated model.
- To learn how to simulate co-occurrence occupancy data.

BACKGROUND READING

The co-occurrence occupancy model is a model option that can be run in the program PRESENCE, and is based upon the material presented in: MacKenzie D.I., L.L. Bailey, and J. D. Nichols. 2004. Investigating species co-occurrence patterns when species are detected imperfectly. Journal of Animal Ecology 73:546-555. This model is described in Chapter 8 of the book, Occupancy Estimation and Modeling. You should have completed the exercise on basic occupancy models before starting this one.

MODEL BACKGROUND

In a simple (or single species, single-season) occupancy model, the goal is to determine the probability that a **site** is occupied by a single species, with the assumption that animals are detected imperfectly. Occupancy models take into account and estimate the probability of species occurrence, based on detection and non-detection data (0 or 1) across sites. One of the benefits of an occupancy model is its ability to estimate species occurrence, even in the case when no animals are detected at a site.

A co-occurrence occupancy model, just like a single species occupancy model, relates encounter histories and detection probabilities to a site. Like a single species occupancy model, a co-occurrence model also assumes that animals are detected imperfectly. The goal of a co-occurrence model, however, is to determine whether a site is occupied by two different species, and to assess if the two species affect each other's detection and occupancy probabilities. Practical problems in ecology include predator-prey interactions and competitive exclusion. A co-occurrence model allows you test whether one species occurs at a site more or less often than is expected if we assume independence of both species. We can also determine if the probability of detection of one species changes in the presence of the other species.

As with the single species, single-season occupancy model, sampling is done during a short period of time when the population at the sites can be considered closed, meaning that the occupancy status of a site does not change between sampling occasions. When a site is sampled multiple times, an encounter history is obtained for both species A and B, which denotes whether species A and/or species B was detected at the site across surveys. We'll come back to this topic a little later.

ENCOUNTER HISTORY BASICS

Click on the worksheet labeled "Co-Occurrence." In our spreadsheet example, each site has two sampling occasions where the presence or absence of each species (A and B) is noted in separate columns. As with the single-season, single-species model, the two sampling occasions may occur at two different points on a site, or the sampling may be done at the same point at the site, but within a short period of time. The encounter histories are written as they are for single species occupancy models, with a (1) used when a species is detected and a (0) when a species is not detected. Because there are two sampling occasions, there are 2^2=4 encounter histories that are possible per species, or $(2^2)^2$ = 16 total combinations of both species. These histories are located in cells C4:C19 for species A, and in cells D4:D19 for species B.

	C	D	E
2	Species A	Species B	
3	History	History	Total
4	11	11	51
5	11	10	18
6	11	01	16
7	11	00	26
8	10	11	26
9	10	10	26
10	10	01	7
11	10	00	20
12	01	11	22
13	01	10	5
14	01	01	26
15	01	00	26
16	00	11	33
17	00	10	25
18	00	01	21
19	00	00	152
20			500

A **site** history of 10 11 (cells C8:D8) means that species A was detected on the first sampling occasion, but not on the second, while species B was detected on both sampling occasion 1 and 2. A site history of 00 10 (cells C17:D17) means that species A was not detected for either sampling occasion 1 or 2, while species B was detected on sampling occasion 1 but not occasion 2. Again, it is important to note that a non-detection of a species at a site does not mean the species was not present, only that the species was not detected. The species could have either been present and not detected, or it may have been absent. Trying to sort this out is a major goal of the analysis.

Let's assume that 500 different sites were sampled twice, and the frequency of the various encounter histories was tallied (as shown on the spreadsheet). The frequencies of each encounter history can be found in cells E4:E19, and we can see in cell E20 that the frequencies sum to 500. The total number of unique histories is given in cell E21. Remember, the basis for our model is the multinomial distribution, and because there are 16 unique history types, there are 15 history probabilities…the 16th history can be derived by subtraction because the sum of the history probabilities must be 1. This also means that we can run a model that estimates up to 15 parameters…otherwise the model will be overparameterized. The naïve occupancy estimate is computed in cell E22 as the probability that a site has neither species A nor species B.

ENCOUNTER HISTORY PROBABILITIES AND CO-OCCURRENCE PARAMETERS

The encounter history probabilities for this model are similar to the encounter histories that we have used in the past for single-species occupancy models. We now need to compute a probability for each of those encounter histories, using the parameters listed in cells F4:F18.

	F	G	H	I
3	Parameter	Estimate?	Betas	MLE
4	ψA	1		0.50000
5	ψB	1		0.50000
6	ψAB	1		0.50000
7	pA1	1		0.50000
8	pA2	1		0.50000
9	pB1	1		0.50000
10	pB2	1		0.50000
11	pAb1	1		0.50000
12	pAb2	1		0.50000
13	pBa1	1		0.50000
14	pBa2	1		0.50000
15	pAB1	1		0.50000
16	pAB2	1		0.50000
17	pab1	0		0.50000
18	pab2	0		0.50000

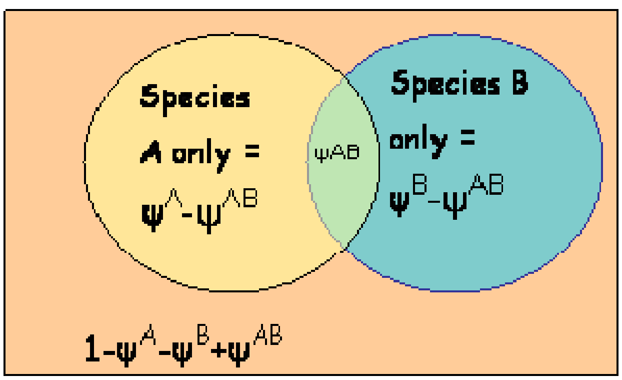
At first it may seem like there are a lot of new parameters being used to estimate encounter history probabilities compared to the other occupancy models. You're right! In fact, these models are very data-hungry; they require a lot of data to adequately estimate parameters. First, each of these parameters is named on the spreadsheet so that the name will appear in formulas rather than a cell reference. For example, click on cell I4 and you should see the name psiA appear to the left of the formula bar. Click on cell I5 and you should see the name psiB appear to the left of the formula bar. Take some time now to get familiar with the names we used for the various parameters in the spreadsheet by clicking on cells I4:I18. MacKenzie *et al.* (2004) provide the table below to define these parameters (modified only slightly to fit our spreadsheet needs), and we'll go through them one by one.

ψA	Probability of species A being present at site i, regardless of occupancy status of species B
ψB	Probability of species B being present at site i, regardless of occupancy status of species A
ψAB	Probability of both species being present at site i.
pA1	Probability of detecting species A during the first survey, given only A is present.
pA2	Probability of detecting species A during the second survey, given only A is present.
pB1	Probability of detecting species B during the first survey, given only B is present.
pB2	Probability of detecting species B during the second survey, given only B is present.
pAb1	Probability of detecting only species A during the first survey, given A and B are present.
pAb2	Probability of detecting only species A during the second survey, given A and B are present.
pBa1	Probability of detecting only species B during the first survey, given A and B are present.
pBa2	Probability of detecting only species B during the second survey, given A and B are present.
pAB1	Probability of detecting both A and B during the first survey, given A and B are present.
pAB2	Probability of detecting both A and B during the second survey, given A and B are present.
pab1	Probability of detecting neither A and B during the first survey, given A and B are present.
pab2	Probability of detecting neither A and B during the second survey, given A and B are present.

OCCUPANCY PARAMETERS

The first three parameters (F4:F6) are called occupancy parameters, and define the probability that a site was occupied by species A (ψA), species B (ψB), or both (ψAB). The remaining parameters are called detection parameters.

First, we'll go into a little background that might help to understand when and how to use these 3 occupancy parameters. Because we are considering the presence of two species in our example study, there are four possible states that describe a site during each sampling occasion, as shown in the Venn diagram below: (1) the presence of species A and species B, denoted ψAB (shaded green in the Venn diagram below), (2) the presence of species A only, denoted ψA-ψAB (shaded yellow in the Venn diagram below), 3) the presence of species B only, denoted ψB-ψAB (shaded blue), and 4) the absence of both species A and species B, denoted 1-ψA-ψB+ψAB (shaded orange). These four states are important to understand, and they must sum to 1 (MacKenzie *et al.* Equation 1).



There are 3 psi parameters that are used in a two species co-occurrence model, given in cells F4:F6. 1) ψ^A, or the probability that species A is present at location *i*, **regardless of the occupancy status of species B**. In the Venn diagram above, this would be the FULL yellow circle, which includes that portion where B also occurs. 2) ψ^B, or the probability that species B is present at location *i*, **regardless of the occupancy status of species A**. In the Venn diagram above, this would be the FULL blue circle, which includes that portion where A occurs. 3) ψ^{AB}, or the probability that **both** species A and species B being present at location *i*. In the Venn diagram above, this would be the intersection of the blue and yellow circles (green). Now, if the two species occurrence were

COMPLETELY independent of each other, we could derive ψAB: $\psi AB = \psi A * \psi B$. In this case, you would estimate ψA and ψB but could then derive ψAB. But if the two species INTERACT with each other, then this might not be true, and you'd have to estimate all three parameters separately. Biologically, this is where things get interesting because understanding the species interactions is the basis of many ecological studies (e.g., predator-prey, competition, parasitism, etc). If the two species are interacting, we are able to determine how the species are interacting by calculating the Species Interaction Factor (SIF), which for occurrence is represented by gamma, γ. Gamma is estimated from a model where all occurrence parameter terms (ψA, ψB and ψAB) are estimated, and $\gamma = \psi AB/(\psi A * \psi B)$. Gamma is computed for you in cell I20. A gamma term that is less than 1 ($\gamma < 1$) would suggest that there is species **avoidance** (species A and species B appear to co-occur less frequently than would be expected under the assumption of independence). A gamma term that is greater than 1 ($\gamma > 1$) would suggest that there is species **convergence**, or the two species co-occur more frequently than expected under independence. Also note that a gamma term that is approximately equal to 1 ($\gamma \cong 1$) would support an assumption of independence of species occurrence.

OK, there is one last state possible in the Venn diagram. By definition, the probability that the site is not occupied by <u>either</u> species is $1-\psi A-\psi B+\psi AB$. There are two websites that may better help you to understand the reasoning and logic behind Venn diagrams. For a more formal and mathematical view of Venn diagrams, you can visit the Journal of Combinatorics website at: http://www.combinatorics.org/Surveys/ds5/VennEJC.html. For a more basic background with an excerpt from *The Mathematical Universe* by William Dunham, you can visit: http://www.cut-the-knot.org/LewisCarroll/dunham.shtml.

These three occupancy parameters (ψA, ψB, ψAB) are used to describe the four different states of occupancy. It is important here to make sure that you know which psi parameter is used when, so we will go into the how and when of each psi parameter in more depth now. As we go through various states below, keep in mind that occupancy describes one of the four states mentioned.

State 1: Both species are present (ψAB; green area in Venn Diagram). When both species are detected at least once at each site, it is known that both species A and B occur at that site. It doesn't matter if both species are detected at the same time or during different sampling occasions. For example, the histories 11 11, 11 10, and 01 10 are examples where both species were detected on the site, indicating the site was known to be occupied by both A and B. In encounter histories where it is known that both species occupy location *i,* only psi AB (ψAB) is used to in the encounter history probability.

State 2: Only species A is present (yellow area in Venn Diagram). At first you might think the probability of only species A being present at a site would simply be ψA but looking at the Venn diagram, we see that the entire yellow circle represents the presence of species A. The ψA term refers to **ANY** situation where species A is present, including the green section that represents where species A and species B are both present on a site. Therefore, the probability of the presence of species A is the yellow portion of the entire circle plus the green section of the circle, since species A is present in both of those situations. In order to calculate the probability of the presence of only species A, the ψAB term (or the green section of the diagram) must be subtracted from ψA, so that just the yellow portion of the circle remains. Therefore, the probability for state 2, or the presence of only species A, is equal to ψA (the total presence of species A, or the probability of the presence of species A, regardless of the occupancy status of species

B) minus ψAB (the probability of the presence of both species A and B), or computationally - the presence of only species A=ψA-ψAB.

State 3: Only species B is present. Using the same logic for State 2, the probability of the presence of species B is the presence of B (psi B) minus the presence of both species A and B (psi AB), or the presence of only species B=ψB - ψAB.

State 4. Neither A or B is present. If the two circles represent the probability of species A occurring (ψA) and species B occurring (ψB), then any area in the Venn diagram that is "not circle" must be the probability that neither species occurs. Since the probability of all four states must sum to 1, we can determine the probability that neither species occurs as 1-ψA-ψB+ψAB. Why add ψAB at the end? Well, we've subtracted this area two times, once in ψA and once in ψB, so we have to add it back in otherwise we double-count this area.

DETECTION PARAMETERS

Now that we have a handle on the four possible states of occupancy, we need to understand the probability that a species is detected or not, given the occupancy state. Let's move on to cells F7:F10: pA1, pA2, pB1 and pB2. These four parameters are the probabilities of detecting the two different species at different times, given that only one species occupies a site. pA1 is the probability of detecting species A during time 1 at location *i*, given that **only** species A is present (ψA-ψAB). pA2 is the probability of detecting species A during the second sampling session, given that **only** species A is present. As you may now guess, pB1 is the probability of detecting species B during time 1 at location *i*, given that **only** species B is present (ψB-

	F	G
7	pA1	1
8	pA2	1
9	pB1	1
10	pB2	1

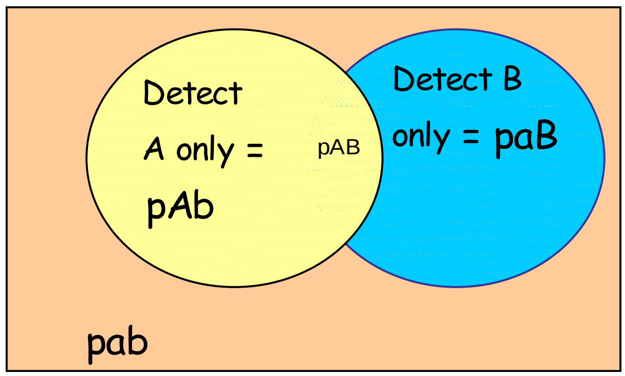
ψAB) and pB2 is the probability of detecting species B at time 2. Because pA or pB is the probability of detecting either species A or B respectively, the probability of NOT seeing species A when it occurs alone is 1-pA, and the probability of NOT seeing species B when it occurs alone is 1-pB.

We'll now move onto the next 8 parameters, listed in cells F11:F18.

	F	G
11	pAb1	1
12	pAb2	1
13	pBa1	1
14	pBa2	1
15	pAB1	1
16	pAB2	1
17	pab1	0
18	pab2	0

These parameters are all various combinations of detecting species A and B at a site, <u>given that both species are present (ψAB)</u>. These parameters include both letters A and B in them (in contrast to pA and pB). For these parameters, size does matter: the size of the letter indicates whether that species was detected or not, *given that both species occurred on the site*. An upper-case A means that species A is detected, and an upper-case B means that species B is detected, while a lower-case a or b means that the species was not detected, although we know that was at the site. Again, for all 8 of these parameters, it is given that both species A and B are present at location *i* during the sampling occasion. So the parameter pAB1 is the probability that both species A and B are detected at location *i* during sampling occasion 1, given both species are present on the site. Parameter paB2 is the probability of detecting species B, but not A at location *i* during the second sampling occasion, given that **both** species are present. Let's try one more to make sure you understand. Parameter pab1 is the probability that neither species A or B are detected at location *i* during the first sampling occasion, given that both are present.

Now, here's a quick quiz. If you know pAB1 = 0.2, pAb1 = 0.3, and paB1 = 0.4, what must pab1 be? Another Venn diagram might be helpful:



In this Venn diagram, four different outcomes can occur when both species are known to be present on a site: 1) only A is detected, or pAb (yellow), 2) only B is detected, paB (blue), 3) both and A and B are detected, or pAB (green), and 4) neither A or B is detected, pab (orange). Note that paB and pAb are already defined as the unique areas in the Venn diagram in which only one species is detected (we don't need to subtract off pAB). Because the sum of these probabilities must add to 1 (and there are no other options available), we can derive pab as 1-pAb-paB-pAB. That's why there is a 0 entered in cells G17 and G18 and these parameters are not directly estimated; you can derive them by subtraction.

CREATING ENCOUNTER HISTORY PROBABILITIES WITH OCCUPANCY AND DETECTION PARAMETERS

Now after all that talk about parameters and Venn diagrams, we can get back to the original issue at hand, namely to estimate the probability of observing each encounter history with the different co-occurrence parameters in the model. After completing all sampling occasions for all sites (two sampling occasions at 500 sites for our spreadsheet example), the frequencies for all the encounter history probabilities are tallied. As explained in the single species, single season occurrence model exercise, each encounter history has a probability of being observed depending on the different detection and occupancy parameters (cells F4:F18). Let's go through a few examples of how the encounter history probabilities are calculated.

For an encounter history of 11 11 (both species A and B are detected during both of the sampling occasions), the probability is pretty straightforward. We find that it is easier to calculate these probabilities by first considering the psi, or probability of presence, and then going through all the different possible detection scenarios for each sampling occasion. So for a history of 11 11, we know that both species are present at the site, so we use psi AB (ψAB). Because we know that both species are present, we know that we will **NOT** use any of the pA or pB terms for the species detection probabilities; rather we will use a combination of the pAB terms because both species are known to occur at the site. We also know that both species were present during both sampling occasions, so we will use those detection probabilities with upper-case letters for both sampling occasions. This results in an encounter history probability of psiAB*pAB1*pAB2, and this history can be seen in cell J4 (=psiAB*p_AB1*p_AB2).

Let's try another encounter history, 10 01. Again we will use ψAB, since both species A and B are known to be present at the site. Remember that it is not important in which sampling occasion the species was detected, just that the species was detected at some time during the sampling. For the first sampling occasion, species A was detected and species B was not, while species B was detected in sampling occasion 2, but species A was not. Again, because both species occur together, we will use a combination of the pAB terms and the size of the letters indicates which species was detected. In the first occasion, A was detected but B was not, so we will use pAb1. In the second occasion, B was detected but A was not, so we will use paB2, which results in an encounter history probability of psiAB*pAb1*paB2, and this history can be seen in cell J10 (=psiAB*pAb1*pBa2). (Note: we 'named' cell paB2 as pBa2, because the "naming cells" option in Excel is insensitive to case. In any event, an upper case indicates that the species was detected, while a lower case indicates the species was not detected, regardless of order).

Now that you are hopefully getting the hang of this, we'll try a more complex encounter history, one in which one of the species was not detected at all, like 00 10 (cell J17). In this history, species A was not detected and species B was detected only during the first sampling occasion. An encounter history where one species is not detected has two different possibilities: either that the undetected species was present but not detected, OR that the undetected species was truly not present at the site. Again, because we cannot be sure if a non-detection in an encounter history is due to the species not being present or simply because it was just not detected, we must consider both situations in our encounter history probability.

So, for the first possibility that species A was present but not detected, we use ψAB for the species presence term, because we are assuming that both species are present. We then use paB1 for the first occasion because species A was not detected but species B was. For the second occasion, we use pab2, because neither species A or B was detected at the site. Therefore our first term is ψAB*paB1*pab2.

The second possibility is that the site was only occupied by species B. Referring back to our Venn diagram and long discussion about how the presence parameters are derived, we (hopefully!) remember that when only species B occurs, it is the **total** presence of species B, psi B (ψB), minus the presence of both species A and B, psi AB (ψAB). Therefore, the probability that only species B was present at the site is ψB-ψAB. Hopefully you also remember that whenever the site is occupied by a single species, we no longer use the pAB terms, but rather the pA or pB terms. In this case, because it was only species B that was detected, we use only pB. For our example encounter history of 00 10, species B was detected during the first occasion, but not during the second occasion, so we use pB1 and (1-pB2), which results in the second part of the encounter history equal to (ψB-ψAB)*pB1*(1-pB2). When combined with the first part of the encounter history, we get: ψAB*paB1*pab2 + (ψB-ψAB)*pB1*(1-pB2). This probability can be seen in cell J17 as =psiAB*pBa1*pab_2+(psiB-psiAB)*pB1*(1-pB2).

To make sure that you understand how each of the encounter history probabilities is calculated, go through each of the probabilities in cells M18:M33. It is important to not skip this step because it is critical to understand how the history probabilities are computed!

The natural log of each history probability is computed in cells K4:K19 for the log likelihood calculation, discussed next.

THE CO-OCCURRENCE LOG LIKELIHOOD

Ultimately, we are going to estimate the occupancy and detection parameters, which will determine the probability of realizing each history, which in turn determines the log likelihood. Our goal is to find those combinations of beta values that maximize the likelihood. We've seen this before in the other spreadsheet exercises:

Betas → MLE's → History probabilities → Log Likelihood.

But first, clear out any betas in cells H4:H18, and you'll see that the sum of the history probabilities is not 1! (It's 2.5). We know that this sum must be 1, so, as we mentioned earlier, we'll need to add two constraints when we run Solver:

1. pab1 = 1-pAb1-paB1-pAB1
2. pab2 = 1-pAb2-paB2-pAB2

We'll add these constraints when we run models, and you will see that the sum of the histories equals 1.

Now that we have our encounter probabilities set up in cells J4:J19, it is time to estimate the values for the different parameters. In this respect, the co-occupancy model is very similar to a single species occupancy model. Again, for a basic background on the mechanics of finding maximum likelihood estimators (MLE's) and the other aspects of deriving estimates for selected parameters, refer to the spreadsheet instructions for single-species occupancy models (Exercise 3). Below is the maximum log likelihood equation, which is the function that we will be trying to maximize. In our example, there

$$\ln(L(p_i \mid n_i, y_i) \propto y_1 \ln(p_1) + y_2 \ln(p_2) + y_3 \ln(p_3) + \ldots.. + y_{16} \ln(p_{16})$$

are 16 different encounter histories (y_i's) and history probabilities (p_i's).

MODEL OUTPUT

The key model outputs are given in cells C24:M25.

	C	D	E	F	G	H
24	Effective N	Log_eL	$-2Log_eL$	AIC	AICc	K
25	500	-767.73	1535.451527	1561.4515	1562.200498	13

	I	J	K	L	M
24	Model DF	Deviance	c-hat	Chi-Sq	Chi-Sq P
25	3	-786.1963242	-262.0654414	501.54	2.2136E-108

The results above reflect a model where all the MLE's = 0.5 – this isn't a valid model because the sum of the encounter histories is 2.5, but for now we just want to point out the key outputs and how they are computed. The effective sample size is the number of sites, and is computed in cell C25. The Log_eL is computed in cell D25, and follows the formula depicted above. It should look very familiar to you by now. The equation in cell D25 is =SUMPRODUCT(E4:E19,K4:K19). The

$-2Log_eL$ is computed in cell E25, and AIC and AICc are computed in cells F25:G25 in the same way as the previous exercises. K, the number of unique parameters estimated, is computed in cell H25 with the formula =SUM(G4:G16). Note again that we won't count pab1 and pab2 as parameters because these can be estimated from the other parameters.

THE MODEL CHI-SQUARE

	L	M
2	CHI-SQUARE	
3	Expected	Chi-Square
4	62.5	0.00
5	62.5	40.00
6	62.5	40.00
7	62.5	20.16
8	62.5	22.50
9	62.5	22.50
10	62.5	52.90
11	62.5	35.34
12	62.5	22.50
13	62.5	52.90
14	62.5	22.50
15	62.5	35.34
16	62.5	12.54
17	62.5	28.22
18	62.5	28.22
19	312.5	65.90
20	1250	501.54

The remaining output gives information on the model degrees of freedom, its deviance, and chi-square results. These are computed in the usual way as well, but let's run through them quickly. The Chi-Square results are computed in cells L4:M19, with the expected values computed in column L. Remember, we obtain the expected value by simply multiplying the probability of each history by the total number of sites. The results shown depict the expected values when the betas are cleared, so the MLE's are all 0.5. This is a bad (and invalid) model! The chi-square value for each history is computed as $(\text{Observed-Expected})^2/\text{Expected}$. For instance, the formula in cell M4 is =(E4-L4)^2/L4. The total chi-square result is the sum of cells M4:M19, and is reported in cell M20 and again in cell L25. The probability of observing a chi-square value this large, given the model's degrees of freedom, is computed in cell M25. Remember that the model's degree of freedom is the number of encounter histories (16) minus K.

THE SATURATED MODEL

	N	O
2	SATURATED MODEL	
3	Probability	Ln Probability
4	0.102	-2.28278247
5	0.036	-3.32423634
6	0.032	-3.44201938
7	0.052	-2.95651156
8	0.052	-2.95651156
9	0.052	-2.95651156
10	0.014	-4.26869795
11	0.04	-3.21887582
12	0.044	-3.12356565
13	0.01	-4.60517019
14	0.052	-2.95651156
15	0.052	-2.95651156
16	0.066	-2.71810054
17	0.05	-2.99573227
18	0.042	-3.17008566
19	0.304	-1.19072758
20	Log_eL =	-1213.83269
21	$-2Log_eL$ =	2427.66537

The saturated model is computed as we've done in the past….simply compute the proportion of each history directly from the raw data, take the natural logs of each proportion, and add them up. These proportions are taken directly from the raw data, so none of the co-occurrence parameters are involved. The proportions are calculated by dividing the observed frequencies by the total number of sites. For example, the formula in cell N4 is =E4/$C$25, and computes the proportion of 11 11 histories in the study. The natural log of this proportion is calculated in cell 04. The saturated model's Log_eL is computed in cell O20, and the $-2Log_eL$ is computed in cell O21. Remember, this is the "standard" upon which all of our co-occurrence occupancy models will be judged. You can run models to your heart's content, but you will not find a $-2Log_eL$ that is greater than the saturated models $-2Log_eL$. Deviance is computed in cell J25, and is the difference between the co-occurrence model's $-2Log_eL$ and the saturated model's $-2Log_eL$. The lower the co-occurrence model's deviance, the closer it is to depicting the actual, observed field data.

RUNNING MODELS

In this exercise, we'll be running 3 models. After running our models in the spreadsheet, we can then compare our answers to the PRESENCE outputs for the same models. One thing to note is that our spreadsheet is set up to estimate ψA, ψB and ψAB, from which we are able to estimate gamma with the equation $\gamma = \psi AB/(\psi A \psi B)$. PRESENCE, however estimates gamma and then derives ψAB from the ψA, ψB and gamma terms. Because you can derive ψAB from γ, and can derive γ from ψAB, ψA, ψB, the estimates from the spreadsheet and the PRESENCE output should be equivalent.

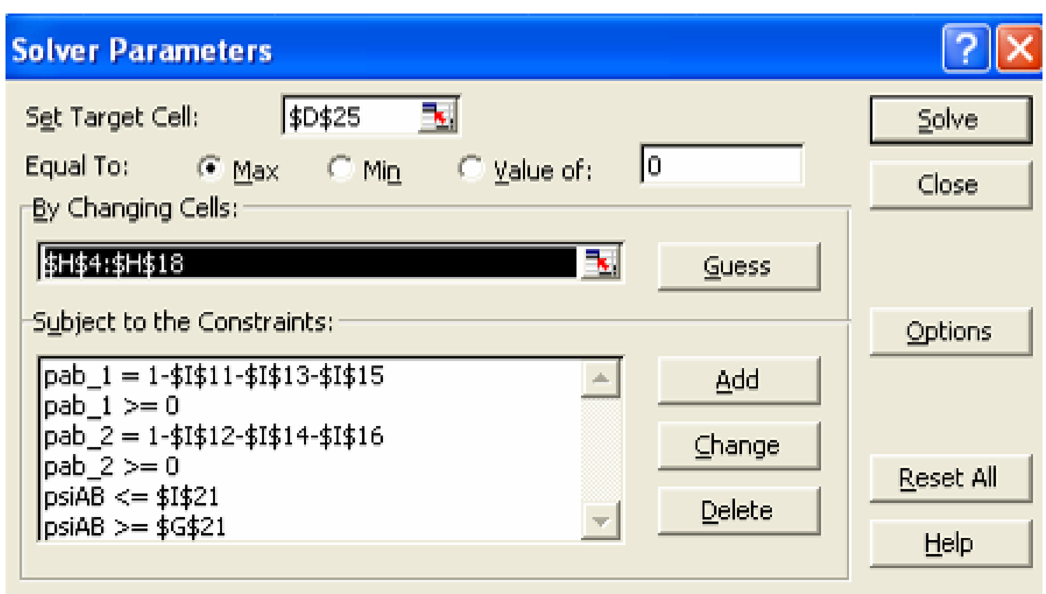
MODEL 1. MODEL PSIA, PSIB, PSIAB, PA (T) PB (T) PAB(T)

For this fully time independent model, we will estimate all of the thirteen parameters for a two species co-occupancy model. Enter a 1 next to each parameter that will be estimated, and clear all of the betas next to those parameters. Enter a 0 in cells G17:G18, because these parameters can be derived from the other parameters (and we don't want to count them as parameters per se, although we'll let Solver work on their betas).

	F	G	H	I
3	Parameter	Estimate?	Betas	MLE
4	ψA	1		0.50000
5	ψB	1		0.50000
6	ψAB	1		0.50000
7	pA1	1		0.50000
8	pA2	1		0.50000
9	pB1	1		0.50000
10	pB2	1		0.50000
11	pAb1	1		0.50000
12	pAb2	1		0.50000
13	pBa1	1		0.50000
14	pBa2	1		0.50000
15	pAB1	1		0.50000
16	pAB2	1		0.50000
17	pab1	0		0.50000
18	pab2	0		0.50000
19	ϕ			1
20	γ			2
21	max (φA+φB-1,0)	0	min (φA, φb)	0.50000

For this example, the maximum likelihood estimates of these parameters are calculated in cells I4:I18 by changing the betas in cells H4:H18. The betas and actual MLE's are linked together with a logit link, which takes the form of $[e^{\beta}/(1+e^{\beta})]$ and can be seen in cells I4:I18. For a review of the logit link and how it is set up, refer to Exercises 4 and 5 for a single species occupancy model with covariates.

Now, open Solver. For this analysis, we want to maximize the log likelihood (cell D25) by changing cells H4:H18. We mentioned that we needed to add constraints on pab1 and pab2. First, we know that they can't be negative because they are probabilities, so enter the constraint that pab1 and pab2 are greater than or equal to 0. Next, we know that pab1 and pab2 can be derived from the other detection parameters, so add the constraints that pab1 = 1-pAB1-pAb1-paB1 and pab2 = 1-pAB2-pAb2-paB2 (if it's not already present). There are two additional constraints that we haven't yet mentioned: psiAB must be less than cell I21 and must be greater than cell G21. Why these constraints? These are logical constraints on the minimum and maximum overlap that is possible, given psiA and psiB. For example, if psiA = 0.6 and psiB = 0.6, then psiAB must be at least 0.2, or a 20% overlap. If there was complete overlap between the two species, then psiAB would be 0.6. A constraint is needed to force ψAB to be between 0.2 and 0.6….values not within this range are not valid. Hopefully this makes sense.



Click Solve and Solver will find those combinations of betas that maximize the log likelihood. It might take a while for Solver to crank through the analysis. You should get the following results, and we'll compare these results to PRESENCE. Here are the model outputs and parameter estimates:

	F	G	H	I
3	Parameter	Estimate?	Betas	MLE
4	ψA	1	0.617165689	0.64957
5	ψB	1	0.530946668	0.62970
6	ψAB	1	0.103698998	0.52590
7	pA1	1	0.168674717	0.54207
8	pA2	1	0.535529866	0.63077
9	pB1	1	1.003587756	0.73176
10	pB2	1	0.438057576	0.60780
11	pAb1	1	-1.858057545	0.13493
12	pAb2	1	-1.841404312	0.13689
13	pBa1	1	-1.524388183	0.17882
14	pBa2	1	-1.319556474	0.21089
15	pAB1	1	-0.159724113	0.46015
16	pAB2	1	-0.251931819	0.43735
17	pab1	0	-1.23046247	0.22610
18	pab2	0	-1.295789169	0.21487
19	ϕ			1
20	γ			1.285700086

	C	D	E	F	G	H
24	Effective N	Log_eL	$-2Log_eL$	AIC	AICc	K
25	500	-1214.18	2428.358205	2454.3582	2455.107176	13

These results suggest that there was a ~0.65 probability that species A would occur on the sites, and a ~0.63 probability that species B would occur. Because ψAB = 0.526, it indicates that the 52.6% of sites had both species. Note for this model that ψA*ψB does not equal ψAB! If the two species interacted completely independently of each other, ψAB would equal 0.64598 * 0.62972 = 0.409, and gamma would equal 1. Solver found that ψAB = 0.526, and gamma = 1.286. Is this strong evidence for positive species interactions, or could it be due to sampling error?

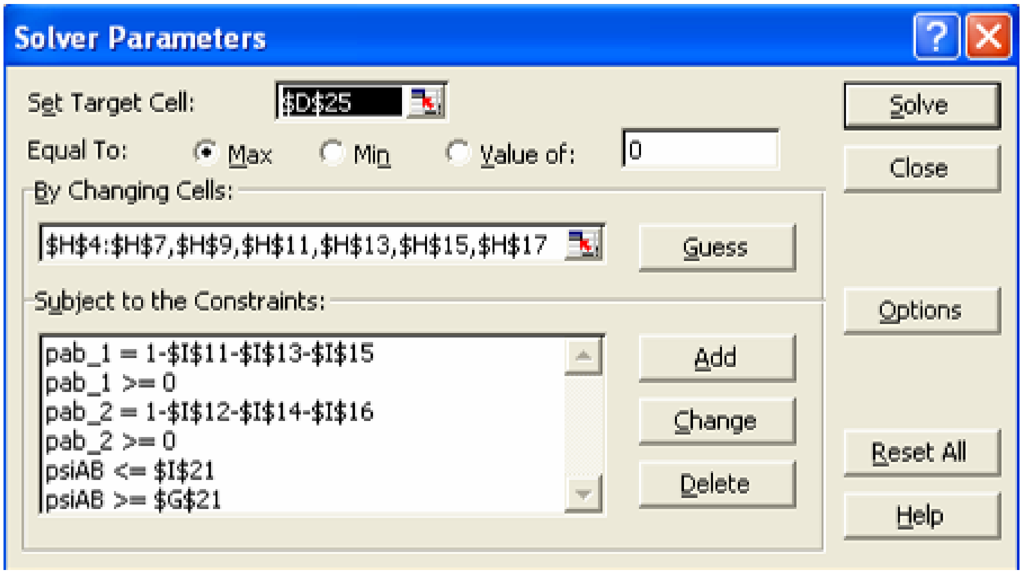
In their 2004 paper, MacKenzie *et al.* indicated that if a species occurs at a site independently of the other, ψAB should equal $\psi A*\psi B$ (which we talked about before in the estimation of gamma). To test whether the species are occurring independently or if there is some interaction, we can set up a likelihood ratio test (LRT). In an LRT, we compare the $-2Log_eL$ from a model which estimates ψA, ψB and ψAB separately and the $-2Log_eL$ from a reduced model that only estimate ψB and ψA (where all other estimated parameters would be the same between the two models). In the second model, a constraint is added so that $\psi AB = \psi A*\psi B$ (independence). We then find the difference in the $-2Log_eL$ for these two models to determine if there is a significant difference between the two models. This difference is Chi-Square distributed, and so if the difference in $-2Log_eL$'s is great, it suggests "significant" evidence for species interactions (either positive or negative); gamma can then be used to determine the magnitude of interaction. Let's try comparing two models in this way, but we'll first constrain the detection parameters to be constant over sampling periods (to keep things simple).

MODEL 2. MODEL PSIA, PSIB, PSIAB, DETECTION (.)

	F	G	H
3	Parameter	Estimate?	Betas
4	ψA	1	
5	ψB	1	
6	ψAB	1	
7	pA1	1	
8	pA2	0	=H7
9	pB1	1	
10	pB2	0	=H9
11	pAb1	1	
12	pAb2	0	=H11
13	pBa1	1	
14	pBa2	0	=H13
15	pAB1	1	
16	pAB2	0	=H15
17	pab1	0	
18	pab2	0	=H17

In this model, we will estimate three psi terms individually, and set the detection constant across time periods. So which of the parameters are we going to estimate? As for the detection parameters, we need a different detection probability for each species, but that detection probability parameter is going to remain constant over time, so we will estimate pA1, pB1, pAB1, pAb1, and paB1, and then will force the betas for session 2 to match the betas for sampling session 1. This cuts down on the number of parameters we need to estimate. So enter a 1 in cells G4:G7, G9, G11, G13, G15, and enter a 0 in the remaining "estimate?" cells. For the sample occasion 2, enter an equation in the beta cells for sample session 2 that forces them to be equal to the corresponding beta for sample occasion 1. Remember that pab2 is also constrained in Solver.

Now, we will maximize cell D25 by changing cells H4:H7,H9,H11,H13,H15, and H17, subject to the usual constraints:



Here are our results:

	F	G	H	I
3	Parameter	Estimate?	Betas	MLE
4	ψA	1	0.621025372	0.65045
5	ψB	1	0.534893152	0.63062
6	ψAB	0	0.104085085	0.52600
7	pA1	1	0.334925466	0.58296
8	pA2	0	0.334925466	0.58296
9	pB1	1	0.675965516	0.66284
10	pB2	0	0.675965516	0.66284
11	pAb1	1	-1.850338548	0.13583
12	pAb2	0	-1.850338548	0.13583
13	pBa1	1	-1.416859479	0.19515
14	pBa2	0	-1.416859479	0.19515
15	pAB1	1	-0.206032595	0.44867
16	pAB2	0	-0.206032595	0.44867
17	pab1	0	-1.26369092	0.22034
18	pab2	0	-1.26369092	0.22034
19	ϕ			1
20	γ			1.282326513
21	max (ψA+ψB-1,0)	0.281075355	min (ψA, ψb)	0.63062

And the model outputs are as follows:

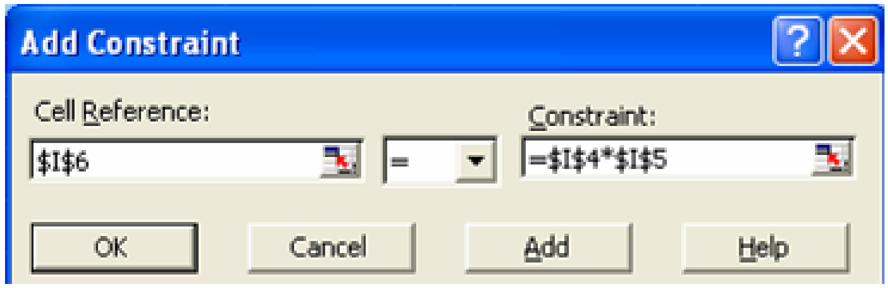
	C	D	E	F	G	H
24	Effective N	Log_eL	$-2Log_eL$	AIC	AICc	K
25	500	-1215.07	2430.137805	2446.1378	2446.431084	8

Now, let's run the exact same model, but this time we will constrain psiAB to be equal to psiA * psiB.

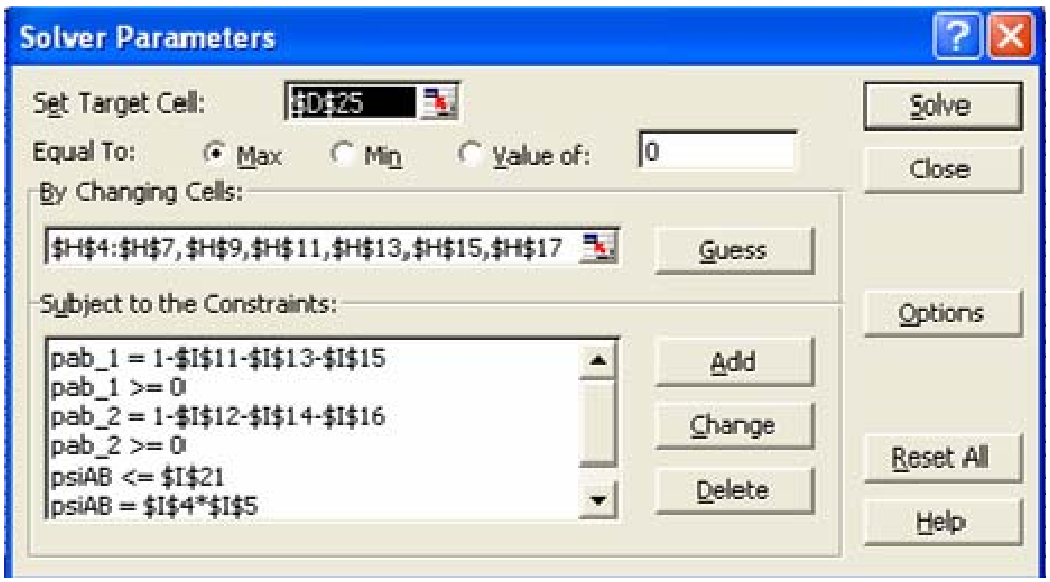
MODEL 3. MODEL PSI A, PSI B, DETECTION(.)

The model that we just ran assumed a different probability of species presence for all 3 psi terms and a different probability of detection for the two different species, although the detection probability does not change with time. For our next model, we will run Solver to estimate all the same parameters except for psi AB. This model could then be considered a reduced model version of model 2, so, if we were going to test whether the species occurred independently, we would be able to use an LRT test on these two models. By doing this, we would be able to determine if the two species occurred independently or not.

For this example, setting up the Betas column is pretty straightforward. It is the same as the last model, except that a 0 should be entered in cell G6 and new constraint must be entered in Solver: namely psiAB = psiA*psiB. Open Solver, and click on the ADD button to add the following constraint:



Click OK. Your Solver dialogue box should now look like this:



Note that you must constrain the MLE's rather than the betas. Why? Because the product of two betas will not necessarily equal the product of two logit transformed betas. Run your model and examine your results:

	F	G	H	I
3	Parameter	Estimate?	Betas	MLE
4	ψA	1	1.345019422	0.79331
5	ψB	1	0.606783411	0.64721
6	ψAB	0	0.053765957	0.51344
7	pA1	1	-1.319403847	0.21092
8	pA2	0	-1.319403847	0.21092
9	pB1	1	0.420364271	0.60357
10	pB2	0	0.420364271	0.60357
11	pAb1	1	-1.615379365	0.16584
12	pAb2	0	-1.615379365	0.16584
13	pBa1	1	-1.509241756	0.18105
14	pBa2	0	-1.509241756	0.18105
15	pAB1	1	-0.176128234	0.45608
16	pAB2	0	-0.176128234	0.45608
17	pab1	0	-1.404997428	0.19702
18	pab2	0	-1.404997428	0.19702
19	φ			1
20	γ			1
21	max (ψA+ψB-1,0)	0.440520878	min (ψA, ψb)	0.64721

	C	D	E	F	G	H
24	Effective N	Log_eL	$-2Log_eL$	AIC	AICc	K
25	500	-1224.06	2448.125845	2462.1258	2462.353487	7

These results are for the model where PsiAB is independent – testing for no species interaction. The difference in the $-2Log_eL$ between the previous model and this model is 2430.14 – 2448.13 = 17.99, which is large and significant. Because the non-independence model had a much lower $-2Log_eL$, we would conclude evidence for species interactions. Gamma from the first model was 1.28. Because it is positive, it suggests evidence of species co-occurrence, rather than avoidance.

RUNNING OTHER MODELS

There are several other models we could run in the spreadsheet. In this exercise, we focused on evaluating whether occurrence of two species was independent or not. We could also run models to determine if detection probability is independent or not. MacKenzie et al. suggest 4 different detection models (their Table 3):

1. pAB = pA * pB.
2. pAB = pA * (1-pB).
3. paB = (1-pA)*pB
4. pab = (1-pA)*(1-pB).

Now that you have a general idea of how to run the models, you could go ahead and run these as well.

SIMULATING CO-OCCURRENCE DATA

	R	S	T	U	V
1	SIMULATE DATA BASED ON EXPECTATION				
2			# Sites =	500	
3			Parameters	Values	
4	11	11	ψA	0.60000	0.1250000
5	11	10	ψB	0.60000	0.0250000
6	11	01	ψAB	0.50000	0.0250000
7	11	00	pA1	0.70000	0.0540000
8	10	11	pA2	0.70000	0.0500000
9	10	10	pB1	0.70000	0.0500000
10	10	01	pB2	0.70000	0.0100000
11	10	00	pAb1	0.10000	0.0310000
12	01	11	pAb2	0.10000	0.0500000
13	01	10	paB1	0.20000	0.0100000
14	01	01	paB2	0.20000	0.0500000
15	01	00	pAB1	0.50000	0.0310000
16	00	11	pAB2	0.50000	0.0690000
17	00	10	pab1	0.20000	0.0410000
18	00	01	pab2	0.20000	0.0410000
19	00	00			0.3380000
20					1.0000000
21			NECESSARY CONSTRAINTS:		
22			pab1 >=	0	
23			pab2 >=	0	
24			ψAB <=	0.60	
25			ψAB >=	0.20	

We're almost finished with the spreadsheet. The last thing to cover in this exercise, before launching PRESENCE, is to understand how to simulate co-occurrence data. As in the other spreadsheet exercises, we'll simulate data in two ways. First, by expectation, and second, with stochasticity. The parameters of interest are listed in cells U4:U18, and the number of sites is given in cell U2. In these cells, you enter the number of sites you wish to simulate, and the parameter values. The parameters shown below are the parameters we used to simulate data in this exercise. Note that cells U17:U18 are calculated, so you only are required to enter values for 13 parameters, plus the number of sites.

Now, PAY ATTENTION! There are some necessary constraints – you can't enter any old values that you want. The first is that pab1 and pab2 must be greater than or equal to 0. Some combinations of parameters can force these values to be negative, so they are invalid. The second is that ψAB must be less than cell U24, and must be greater than cell

U25 for reasons we mentioned previously. If you enter parameter values, and either of these constraints are violated, you need to try different values.

CREATING DATA FROM EXPECTED VALUES

	R	S	T	U	V	W
2			# Sites =	500		
3			Parameters	Values		
4	11	11	ψA	0.60000	0.1250000	63
5	11	10	ψB	0.60000	0.0250000	13
6	11	01	ψAB	0.50000	0.0250000	13
7	11	00	pA1	0.70000	0.0540000	27
8	10	11	pA2	0.70000	0.0500000	25
9	10	10	pB1	0.70000	0.0500000	25
10	10	01	pB2	0.70000	0.0100000	5
11	10	00	pAb1	0.10000	0.0310000	16
12	01	11	pAb2	0.10000	0.0500000	25
13	01	10	paB1	0.20000	0.0100000	5
14	01	01	paB2	0.20000	0.0500000	25
15	01	00	pAB1	0.50000	0.0310000	16
16	00	11	pAB2	0.50000	0.0690000	35
17	00	10	pab1	0.20000	0.0410000	21
18	00	01	pab2	0.20000	0.0410000	21
19	00	00			0.3380000	169
20					1.0000000	500

Given the parameters entered in cells U4:U16, and the number of sites entered in cell U2, the expected number of sites showing each history is computed in cells W4:W19. These are computed by multiplying the probability of each history by the number of sites. The probability of each history is simply the equations we already know and love. For instance, the probability of realizing a 11 11 history is psiAB*pAB1*pAB2. This result, multiplied by the number of sites, gives the number of sites expected to have a 11 11 history, and is computed in cell W4. It's that simple. Check out the other equations and make sure there are no mistakes in there.

CREATING DATA WITH STOCHASTICITY

The second way of creating data is by simulation – i.e., involving some element of stochasticity. This method is carried out in columns Y:AN, and is quite a bit more involved than the first method. In this method, we again use the parameters entered in cells U4:U16 and the number of sites entered in cells U2. The spreadsheet is currently set up to simulate data for 500 sites, but you could easily change that by extending or deleting the formula used.

The method for simulating data first involves a bit of calculations, which are done for you in cells Z1:AG7. Notice the message in row 1: HANDS OFF! These values are computed based in the parameter estimates you entered previously. The blue cells focus on the occupancy parameters, and the orange cells focus on the detection parameters. The important thing to recall right now are those Venn Diagrams we walked through earlier in the exercise.

	Z	AA	AB	AC	AD	AE	AF	AG
1	HANDS OFF! THESE ARE COMPUTED VALUES BASED ON MLE'S ENTERED!							
2	OCCUPANCY STATUS			DETECTION STATUS GIVEN A & B PRESENT				
3			0		Session 1	0	Session 2	0
4	A	0.10000	0.10000	A	0.10000	0.10000	0.10000	0.10000
5	B	0.10000	0.20000	B	0.20000	0.30000	0.20000	0.30000
6	AB	0.50000	0.70000	AB	0.50000	0.80000	0.50000	0.80000
7	0	0.30000	1.00000	0	0.20000	1.00000	0.20000	1.00000

Let's study what these cells are doing, starting with the occupancy cells. In cell AA4, we entered the equation =U4-U6, which is psiA-psiAB. So, the result in this cell is the probability that the occupancy space is occupied by species A alone. Similarly, the probability that the occupancy space is occupied by species B alone is computed in cell AA5 with the equation =U5-U6, which is psiB-psiAB. Cell AA6 gives the proportion of the occupancy space that is occupied by both species, and is simply psiAB. So, you can see that our data were simulated with strong overlap between the two species. The probability that the occupancy space is occupied by neither species is computed in cell AA7 with the formula =1-U4-U5+U6. We already know that the sum of the probabilities must be 1.

What about cells AB3:AB7? These are cumulative probabilities. For example, given the parameters entered previously, the probability of getting A alone is 0.1 (cell AB4). Cell AB5 is the probability of A alone OR B alone, and is 0.2. Cell AB6 is the probability of getting A alone, B alone, or A and B, and is 0.7. Cell AB7 is the probability of getting any of the four states, which by definition must be 1. We're going to use these cumulative probabilities for simulating occupancy data.

	Z	AA	AB
2	**OCCUPANCY STATUS**		
3			0
4	A	0.10000	0.10000
5	B	0.10000	0.20000
6	AB	0.50000	0.70000
7	0	0.30000	1.00000

OK, now let's focus on the detection probabilities. Again, we know that there are four possible states when a site is known to be occupied by BOTH species: pAB (both are detected), pAb (only A is detected), paB (only B is detected), and pab (neither are detected). These four probabilities must sum to 1. Cells AD4:AD7 give the probability of detecting A only in session 1 (cell AD4 =U11), detecting B only in session 1 (cell AD5 =U13), detecting A and B in session 1 (cell

	AC	AD	AE	AF	AG
2	**DETECTION STATUS GIVEN A & B PRESENT**				
3		Session 1	0	Session 2	0
4	A	0.10000	0.10000	0.10000	0.10000
5	B	0.20000	0.30000	0.20000	0.30000
6	AB	0.50000	0.80000	0.50000	0.80000
7	0	0.20000	1.00000	0.20000	1.00000

AD6 =U15), and detecting neither in session 1 (cell AD7 =U17). Cells AE4:AE7 are the cumulative probabilities for session 1, and we'll be using these values to assign encounter histories when a site is known to have both species present. Cells AF4:AG7 essentially do the same thing, only focus on session 2.

Given these data, we can now move to the portion of the spreadsheet where the co-occurrence data are actually simulated.

	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH
1		HANDS OFF! THESE ARE COMPUTED VALUES BASED ON MLE'S ENTERED!								
2		OCCUPANCY STATUS			DETECTION STATUS GIVEN A & B PRESENT					
3				0		Session 1	0	Session 2	0	
4		A	0.10000	0.10000	A	0.10000	0.10000	0.10000	0.10000	
5		B	0.10000	0.20000	B	0.20000	0.30000	0.20000	0.30000	
6		AB	0.50000	0.70000	AB	0.50000	0.80000	0.50000	0.80000	
7		0	0.30000	1.00000	0	0.20000	1.00000	0.20000	1.00000	
8										
9			Session 1	Session 2	Species A			Species B		
10	Site	Occupancy	Detect	Detect	Session 1	Session 2	History	Session 1	Session 2	History
11	1	AB	0	0	0	0	00	0	0	00
12	2	B			0	0	00	1	1	11
13	3	0			0	0	00	0	0	00
14	4	AB	AB	A	1	1	11	1	0	10
15	5	AB	B	AB	0	1	01	1	1	11

The sites are listed from 1 to 500 in column Y. In column Z, we use the blue cells (cells AB4:AB7) to establish which species, if any, are present on the site. Click on cell Z11 and you'll see the formula =LOOKUP(RAND(),$AB$3:$AB$7,$Z$4:$Z$7). This LOOKUP function looks up a random number within an ordered set of values (cells AB3:AB7), and returns the appropriate value in cells Z4:Z7. For instance, site 1 has an occupancy of AB. This means that the random number must have been greater than 0.2, and less than 0.7. We know that 50% of the sites are occupied by both species, so this method assures us that roughly 50% of the random numbers between 0 and 1 will be associated with AB. Make sense?

Now that we know which site is occupied by which species, we now need to determine who was detected on those sites. Click on cell AA11, and you'll see the formula =IF(Z11="AB",LOOKUP(RAND(),$AE$3:$AE$7,$AC$4:$AC$7),""). In column AA, we want to determine which species were detected, only if both species are known to occupy the site. In column AB, we want to determine which species were detected, only if both species are known to occupy the site. The formula in cell AA11 first evaluates if the site is occupied by both species IF(Z11="AB"). If this is true, the next formula is carried out LOOKUP(RAND(),$AE$3:$AE$7,$AC$4:$AC$7). If it's not true, a "" is returned, which looks blank on the spreadsheet. Well, site 1 happens to be occupied by both species. So, which species were detected in sampling session 1? Again we use a LOOKUP function, and look up a random number between 0 and 1 within the ordered cumulative probabilities (cells AA3:AA7), and then return the corresponding result: A, B, AB, or 0, where 0 means neither.

	Y	Z	AA	AB
9			Session 1	Session 2
10	Site	Occupancy	Detect	Detect
11	1	AB	B	AB

For site 1, it was occupied by both A and B, and only B was detected in session 1, but both A and B were detected in session 2. This site should end up with a history of 01 11 (you'll see this is so later).

OK, now let's move onto the next section. In column AC, we enter formula to determine if species A was detected in session 1, in column AD we enter a formula to determine if species A was detected in session 2, in column AF we enter a formula to determine if species B was detected in session 1, and in column AG we enter a formula to determine if species B was detected in session 2.

	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH
9			Session 1	Session 2	**Species A**			**Species B**		
10	Site	Occupancy	Detect	Detect	Session 1	Session 2	History	Session 1	Session 2	History
11	1	AB	B	AB	0	1	01	1	1	11
12	2	A			1	1	11	0	0	00
13	3	0			0	0	00	0	0	00

The formulae in these columns have the same logical structure, so we'll work our way only through column AC. Click on cell AC11, and you'll see the formula =IF(Z11="B",0,IF(Z11=0,0,IF(AND(Z11="A",RAND()<$U$7),1,IF(AND(Z11="AB",AA11="A"),1,IF(AND(Z11="AB",AA11="AB"),1,0))))). Looks rough, but let's step through it. The formula is several IF functions. Remember, an IF function has three basic components: the logical value, value if true, value if false. So the first IF function states IF Z11 = "B", then return a 0, otherwise walk through the next IF function. Obviously, if the site is occupied by only species B, then species A cannot be detected so a 0 is returned. If Z11 does not = B, then the equation moves onto the next IF function: IF Z11=0,0. This is straightforward too. If cell Z11 =0, the site is not occupied by either species, so there is no chance of detecting species A in session 1. What if Z11 does not equal 0? The equation moves onto the next IF function: IF(AND(Z11="A",RAND()<$U$7),1. This one is a tad bit trickier. If cell Z11 = A, it means the site is occupied by only A. If it is occupied by only A AND a random number between 0 and 1 is less than pA1 (cell U7), then species A was detected in session 1, and a 1 is returned. If both of those things are not true, the formula moves onto the last IF function: IF(AND(Z11="AB",AA11="AB"),1,0. This one also has a AND function within it. If cell Z11 = AB AND cell AA11 = AB, then a 1 is returned and species A was detected in session 1, otherwise a 0 is returned.

Whew! These formulae are copied down for all 500 sites, and the same idea is carried out for sampling session 2 and for species B in both sampling sessions.

The data are then summarized and tabulated in cells AK11:AN26. Press F9, the calculated key, and Excel will draw new random numbers, and hence new histories will be generated. Press F9 1000 times and store the frequencies each time, and you will have 1000 different datasets each based on the same parameter estimates. This comes in handy for doing bootstrap analysis.

	AK	AL	AM	AN
10	**SIMULATED HISTORIES**			
11	**11**	**11**	11 11	47
12	**11**	**10**	11 10	11
13	**11**	**01**	11 01	12
14	**11**	**00**	11 00	25
15	**10**	**11**	10 11	22
16	**10**	**10**	10 10	19
17	**10**	**01**	10 01	4
18	**10**	**00**	10 00	21
19	**01**	**11**	01 11	33
20	**01**	**10**	01 10	1
21	**01**	**01**	01 01	32
22	**01**	**00**	01 00	10
23	**00**	**11**	00 11	33
24	**00**	**10**	00 10	22
25	**00**	**01**	00 01	27
26	**00**	**00**	00 00	181
27				500

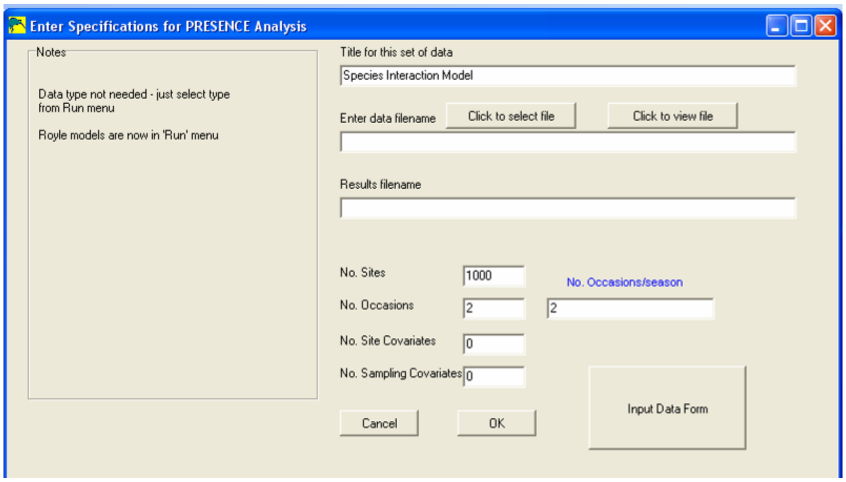
PRESENCE INPUT FILES

OK! We're ready to run some analyses in the program PRESENCE. If you haven't downloaded this program yet, go to http://www.mbr-pwrc.usgs.gov/software/doc/presence/presence.html and download the program. In this program, we can simply paste in our raw data and don't need to create input files. The data used for this exercise are listed in columns AT:AU. We used the old "tally" trick we've used in previous exercise, so won't go into the details here. You'll just copy cells AT2:AU1001 and will paste them directly into PRESENCE.

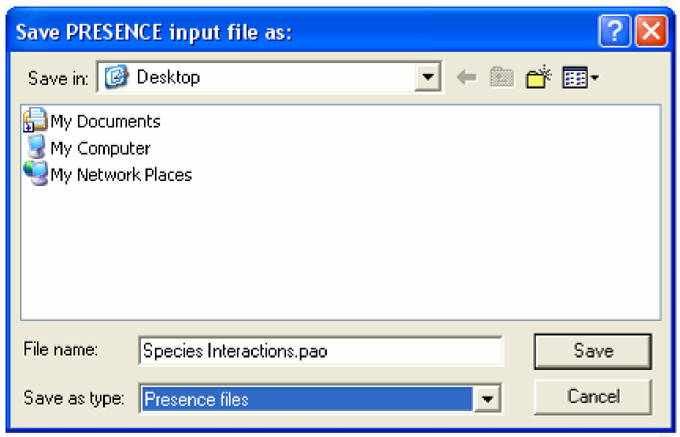
SINGLE-SEASON, SPECIES-INTERACTIONS OCCUPANCY MODELS IN PRESENCE

GETTING STARTED

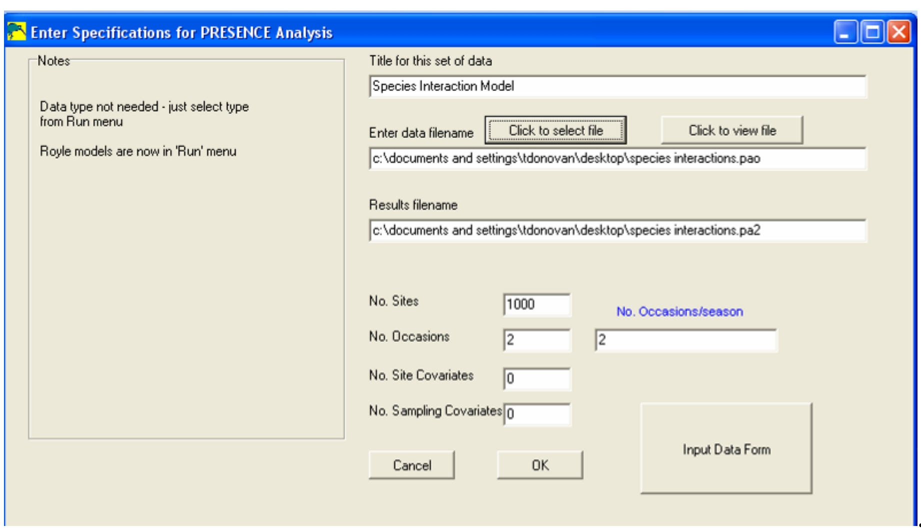
Open PRESENCE and go to File | New Project. A new window will appear. Enter a title for this project (e.g., Species Interaction Model). We had 500 sites for two species, with two sampling occasions. However, with the Species Interactions model, we need to stack the encounter histories on a species-by-species basis. So, instead of entering 500 in the No. Sites textbox, enter 1000. As soon as you select the Species Interactions option, PRESENCE will know that the first 500 encounter histories pertain to species 1, and the next 500 histories will pertain to species 2. Set the No. Occasions/season to 2, and then click on the button labeled "Input Data Form."



A new window that looks similar to an excel spreadsheet will open with two columns for the sampling occasions and 1000 rows for each site (500 for species A and 500 for species B). Go back to the EXCEL workbook. Select the range of Data from the PRESENCE INPUT section of the "Co-Occurance" page (cells AT2:AU1001) and copy. Do not copy any other cells, only the two columns of data. Go back to PRESENCE. Click in the first open cell in the top left. Go to the toolbar and select Edit | Paste | Values. This will fill in the EXCEL data into all the cells. Be sure to save the input data by going to File | Save as, and then enter a name for the input file (e.,g Species Interactions.pao) and click the Save button.



Then, on the Enter Specifications Form, click on the button labeled "Click to select file" and navigate to your freshly created input file:

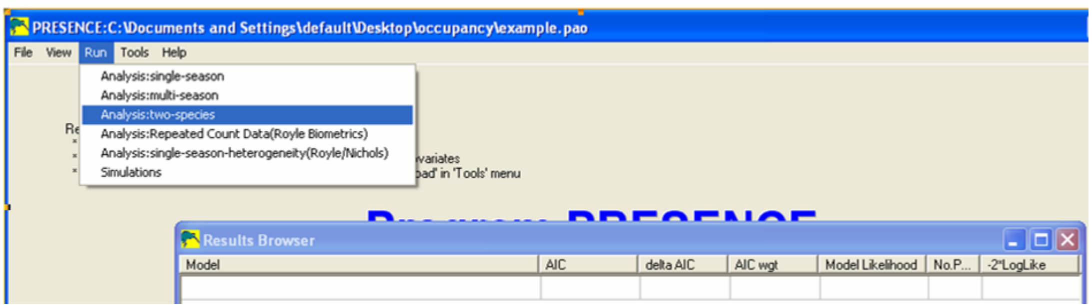


Once things are set, press the OK button.

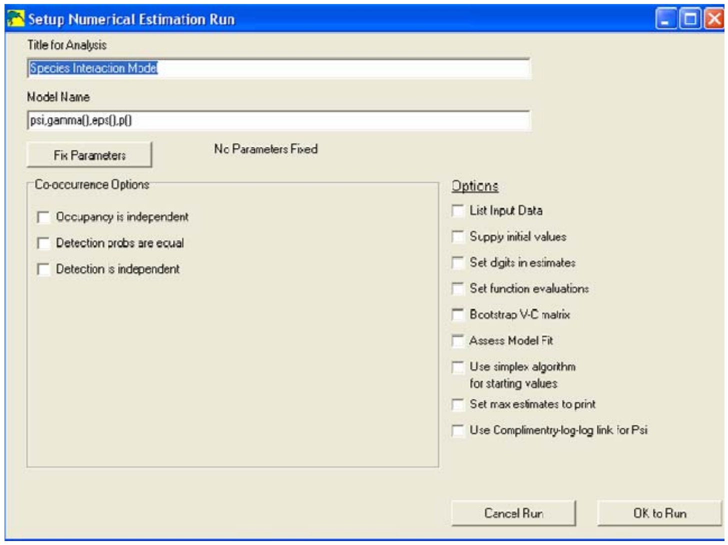
PRESENCE will close the specifications screen and open a Results Browser window.

MODEL 1. MODEL PSIA, PSIB, PSIAB, PA (T) PB (T) PAB(T)

OK, our first model will be where we estimate all parameters uniquely. (That is, we'll estimate ψA, ψB, ψAB, plus survey-specific detection parameters for all parameters except pab1 and pab2, which are derived). To run a model, go to Run | Analysis: two-species:



The next form that appears is the "Setup Numerical Estimation Run" form.

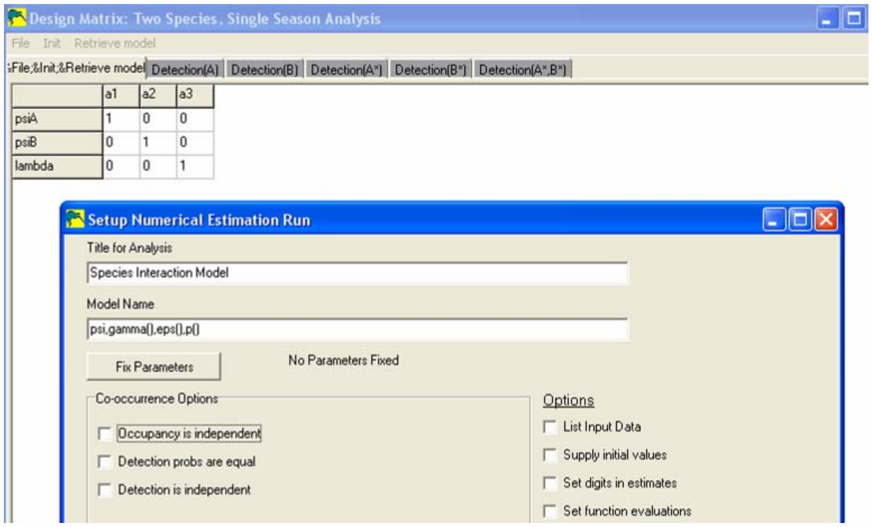


Note that this form has an option box labeled "Co-Occurrence Options." There are three check-boxes:

1. Occupancy is independent
2. Detection probs are equal
3. Detection is independent.

These are not mutually exclusive – you can check multiple boxes, or none. Let's go through what each option means.

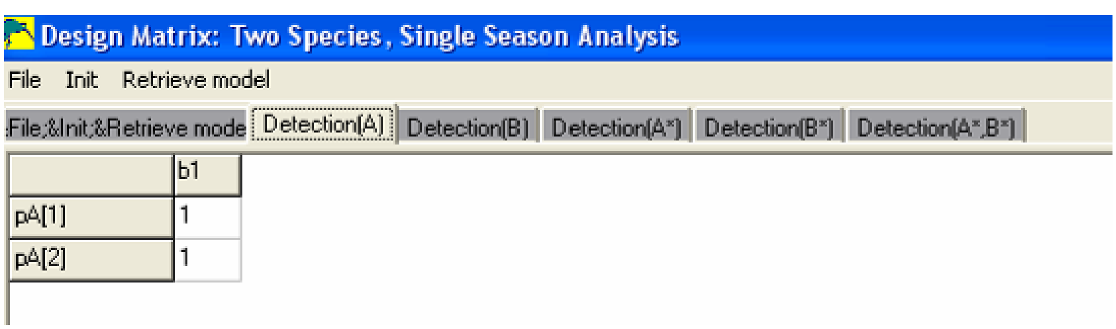
First, make sure that NONE of the boxes are checked, and then look at the Design Matrix (DM):



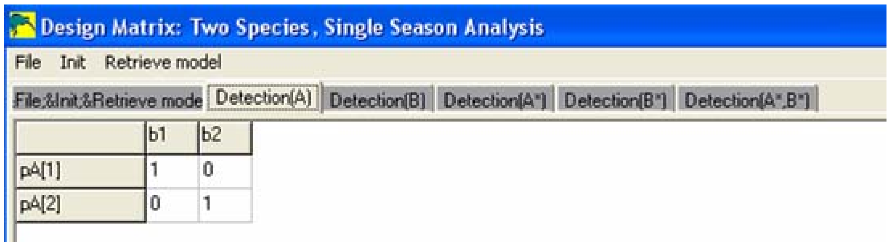
Notice that the DM has 6 tabs, labeled "Occupancy", Detection [A], Detection [B], Detection [A*], Detection [B*], Detection [A*B*].

The Occupancy tab shows estimates for psiA, psiB, and lambda (should this be called gamma?). Remember that our spreadsheet is set up to estimate ψA, ψB and ψAB, from which we are able to estimate gamma with the equation $\gamma = \psi AB/(\psi A\psi B)$. PRESENCE, however estimates gamma and then derives ψAB as the $\gamma * \psi A * \psi B$.

The Detection [A] tab shows estimates for parameters pA1 and pA2 – the probability of detecting species A when only species A occurs on the site. By default, PRESENCE sets pA1 = pA2, and calls the beta estimate that corresponds to these parameters as b1.

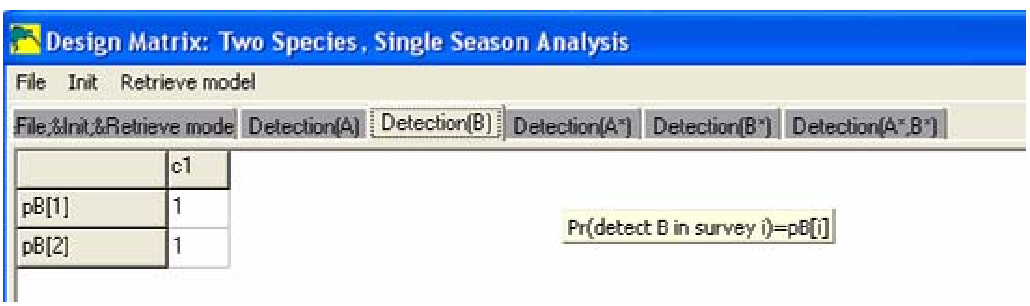


For this first model, we want separate estimates of pA1 and pA2. So click somewhere in the Detection [A] tab, and then select Init | Full Identity:

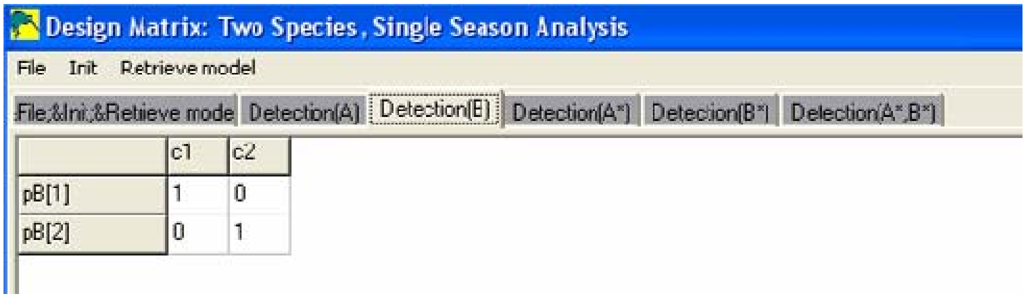


Now you can see that PRESENCE will estimate pA1 uniquely (with a beta called b1), and will estimate pA2 uniquely (with a beta called b2).

Now click on the Detection [B] tab. The Detection [B] tab shows estimates for parameters pB1 and pB2 – the probability of detecting species B when only species B occurs on the site. By default, PRESENCE sets pB1 = pB2, and calls the beta estimate that corresponds to these parameters as c1.

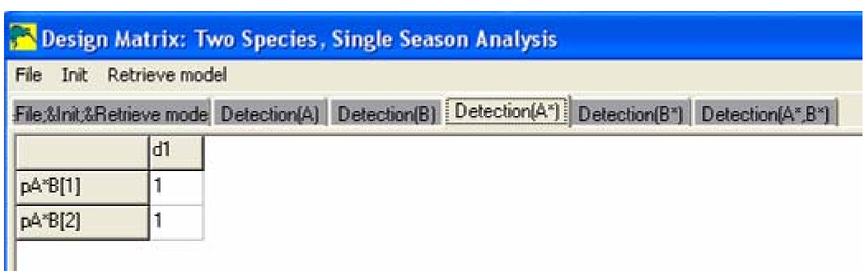


For this first model, we want separate estimates of pB1 and pB2. So click somewhere in the Detection [A] tab, and then select Init | Full Identity:



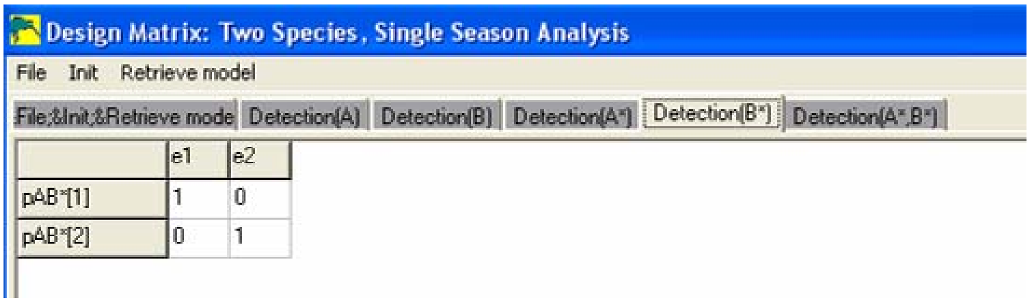
Now you can see that PRESENCE will estimate pB1 uniquely (with a beta called c1), and will estimate pB2 uniquely (with a beta called c2).

The next three tabs are for specifying the detection parameters when both species are present on a site. The Detection [A*] shows estimates for parameters pAb1 and pAb2 – the probability of detecting species A when only when both species occur on the site. By default, PRESENCE sets pAb1 = pAb2, and calls the beta estimate that corresponds to these parameters as d1. (Is there a way to re-label the estimates by getting rid of the B, or at least making it lower case?)



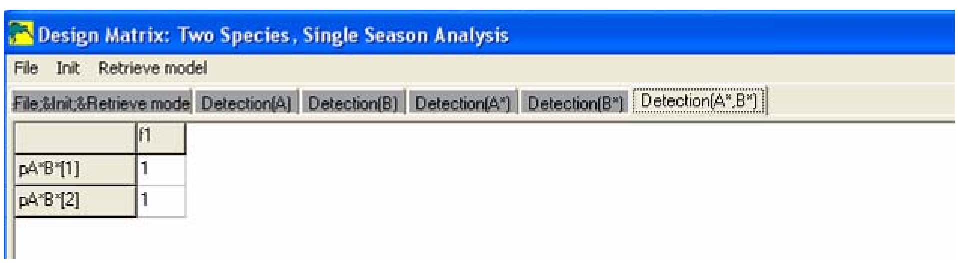
Once again, we need to let PRESENCE know that we want unique estimates for these parameters, so go to Init | Full Identity, and you'll see that the two parameters will be estimated uniquely (with betas defined as d1 and d2).

The Detection [B*] shows estimates for parameters paB1 and paB2 – the probability of detecting species B when only when both species occur on the site. By default, PRESENCE sets paB1 = paB2, and calls the beta estimate that corresponds to these parameters as e1. Again, go to Init | Full Identity to specify that these parameters should be uniquely estimated for this first model:

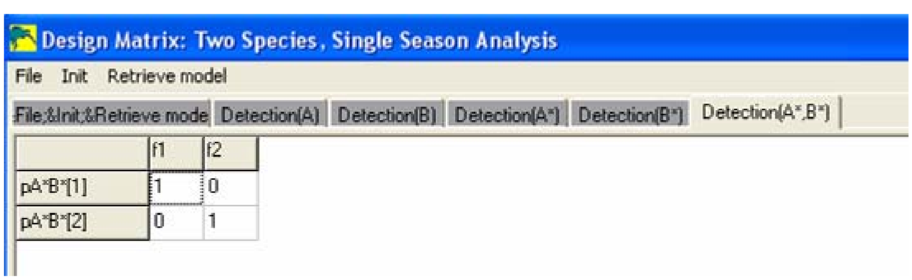


(Is there a way to re-label the estimates by getting rid of the A, or at least making it lower case?)

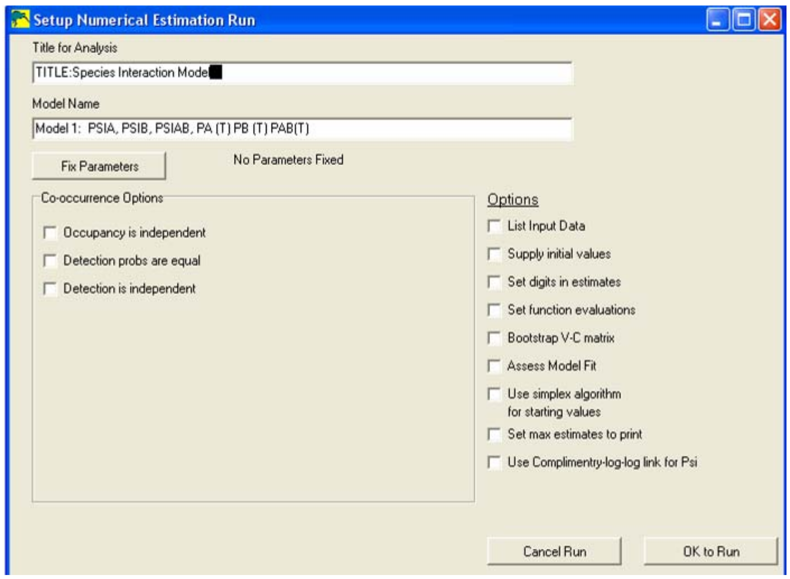
The last tab is labeled Detection [A*B*], which shows estimates for pAB1 and pAB2, or the probability that both species are detected given that both species occur on a site. By default, PRESENCE sets these parameters equal to each other with a beta called f1.



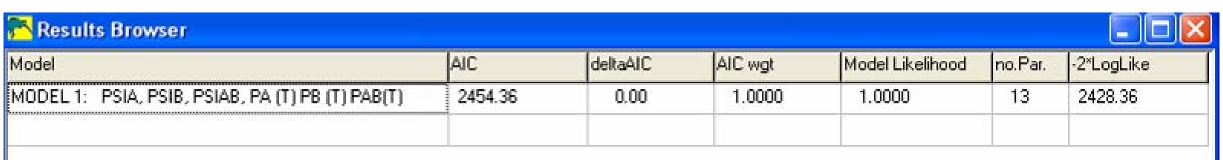
Initialize this matrix by selecting Full Identity.



OK, now that you understand the different tabs in PRESENCE, it should be fairly straight-forward to run other models. We're now ready to run our model: PSIA, PSIB, PSIAB, PA (T) PB (T) PAB(T). Return to the Setup Numerical Estimation Run form, and enter a title for this particular model:



When you are ready, press the "OK to Run" button, and add the results to the Results Browser:


Results Browser

Model	AIC	deltaAIC	AIC wgt	Model Likelihood	no.Par.	-2*LogLike
MODEL 1: PSIA, PSIB, PSIAB, PA (T) PB (T) PAB(T)	2454.36	0.00	1.0000	1.0000	13	2428.36

	C	D	E	F	G	H
24	Effective N	Log_eL	$-2Log_eL$	AIC	AICc	K
25	500	-1214.18	2428.358205	2454.3582	2455.107176	13

You should notice that the $-2Log_eL$ and AIC values match those of the spreadsheet.

Note – I accidentally clicked on the multi-season analysis…and then the DM shows the multi-season tabs rather than the co-occurrence parameters, even though I changed it back to two-species.

Now right-click on the model name to bring up the full model output. Notice that PRESENCE correctly reports the number of sites (500) and number of sampling occasions (2). This model estimated 13 parameters. Now let's look at the beta estimates:

pres4916.tmp - Notepad

File Edit Format View Help

```
Number of sites                  = 500
Number of sampling occasions     = 2
Number of missing observations   = 0

Number of parameters             = 13

Model:model 1:   psia, psib, psiab, pa (t) pb (t) pab(t) flags:(0 0 0 0) (ind-occ det-eq det-ind init)

NPar=13 s=0


**** Numerical convergence was not reached.
     Parameter estimates converged to approximately 5.63 significant digits. ****
-2log(likelihood)                = 2428.358210
AIC                              = 2454.358210
See help file for details on how the model has been fit.
```

So far, so good. Now let's look at the beta estimates.

	F	G	H
3	Parameter	Estimate?	Betas
4	ψA	1	0.617165561
5	ψB	1	0.53094648
6	ψAB	1	0.103698778
7	pA1	1	0.168674629
8	pA2	1	0.535529501
9	pB1	1	1.003587928
10	pB2	1	0.438057084
11	pAb1	1	-1.858057531
12	pAb2	1	-1.841404205
13	pBa1	1	-1.524388324
14	pBa2	1	-1.31955638
15	pAB1	1	-0.159723989
16	pAB2	1	-0.251931328
17	pab1	0	-1.230462537
18	pab2	0	-1.295790052

pres4916.tmp - Notepad

File Edit Format View Help

```
**** Numerical convergence was not reached.
     Parameter estimates converged to approximately 5.63 significant
-2log(likelihood)                = 2428.358210
AIC                              = 2454.358210
See help file for details on how the model has been fit.

Coefficients (SE) for occupancy covariates:
       a1          0.617158 ( 0.132917)   [ 0.649572 ( 0.030256)]
       a2          0.530968 ( 0.124723)   [ 0.629709 ( 0.029082)]
       a3          0.251293 ( 0.153111)   [ 0.562495 ( 0.037680)]
Coefficients (SE) for detection(A) covariates:
       b1          0.168632 ( 0.439293)   [ 0.542059 ( 0.109046)]
       b2          0.535550 ( 0.426986)   [ 0.630777 ( 0.099444)]
Coefficients (SE) for detection(B) covariates:
       c1          1.003469 ( 0.366966)   [ 0.731740 ( 0.072034)]
       c2          0.438007 ( 0.371305)   [ 0.607784 ( 0.088513)]
Coefficients (SE) for detection(AB) covariates:
       d1          0.385085 ( 0.148238)   [ 0.595099 ( 0.035719)]
       d2          0.299154 ( 0.147663)   [ 0.574236 ( 0.036102)]
Coefficients (SE) for detection(AB') covariates:
       e1          0.570933 ( 0.145274)   [ 0.638978 ( 0.033513)]
       e2          0.611289 ( 0.145989)   [ 0.648235 ( 0.033289)]
Coefficients (SE) for detection(A'B) covariates:
       f1          0.190741 ( 0.097902)   [ 0.547541 ( 0.024254)]
       f2          0.161188 ( 0.093049)   [ 0.540210 ( 0.023112)]
```

Let's start with the betas labeled a1, a2, and a3. Remember, these are the betas that PRESENCE labeled for ψA, ψB, and γ. Jim – I need your help here..I have no idea why the betas are so different…this is the problem I raised at the workshop this summer, but I can't remember how it was resolved.

Now let's look at the bottom of the PRESENCE output to view the parameter estimates:

```
------------------------------
Psi is array of occupancy probabilities
   Psi[0] = psi[A] = 0.649572
   Psi[1] = psi[B] = 0.629709
   Psi[2] = psi[AB]= 0.525899

 P is array of detection probabilities
   P[0] = seeing A, B not present = 0.542059
   P[1] = A not present, seeing B = 0.731740
   P[2] = p*A = Pr(detect A | A & B present) = 0.595099
   P[3] = p*B = Pr(detect B | A & B present) = 0.638978
   P[4] = p*AB = Pr(detect A&B | A&B present) = p*A x p*B if independent= 0.460164

CPU time: 3.0 seconds
```

Here it is easier to see the agreement between the spreadsheet and PRESENCE. The occupancy parameters, PsiA, PsiB, and psiAB, estimates match. The detection parameters also match, but PRESENCE output reports the estimates in a summary format. P[0] is the probability of seeing A, given B is absent, and is 0.54206 (which matches cell I7). P[1] is the probability of seeing B, given A is absent, and is 0.7317 (which matches cell I9). P[2] is the probability of detecting A, given both species are present, and is 0.5951 (which is the sum of cells I11 and I15, or the probability of detecting A alone or both A and B). Similarly, P[3] is the probability of detecting B, given both species are present, and is 0.6390 (which is the sum of cells I12 and I16, or the probability of detecting B alone or both A and B). P[4] is the

	F	G	H	I
3	Parameter	Estimate?	Betas	MLE
4	ψA	1	0.617165561	0.64957
5	ψB	1	0.53094648	0.62970
6	ψAB	1	0.103698778	0.52590
7	pA1	1	0.168674629	0.54207
8	pA2	1	0.535529501	0.63077
9	pB1	1	1.003587928	0.73176
10	pB2	1	0.438057084	0.60780
11	pAb1	1	-1.858057531	0.13493
12	pAb2	1	-1.841404205	0.13689
13	pBa1	1	-1.524388324	0.17882
14	pBa2	1	-1.31955638	0.21089
15	pAB1	1	-0.159723989	0.46015
16	pAB2	1	-0.251931328	0.43735
17	pab1	0	-1.230462537	0.22610
18	pab2	0	-1.295790052	0.21487
19	ϕ			1
20	γ			1.285700099

probability of detecting both A and B, given both are present, and is 0.4601 (which matches cell I15 in the spreadsheet). Thus, the results are provided for the first sampling occasion only. (Jim, can this be expanded to include the results for occasion 2?)

MODEL 2. MODEL PSIA, PSIB, PSIAB, DETECTION(.)

MODEL 3. MODEL PSI A, PSI B, DETECTION(.)