EXERCISE 17: INTEGRATED MODELING OF HABITAT SUITABILITY AND OCCUPANCY

In collaboration with Sarah J. Frey[1], William V. DeLuca[2], and Ryan S. Butryn[1]

[1]*Rubenstein School of Environment and Natural Resources, University of Vermont*

[2]*Department of Natural Resources Conservation, University of Massachusetts - Amherst*

OBJECTIVES

- To learn and understand the integrated model of habitat and occupancy, and how it fits into a multinomial maximum likelihood analysis.
- To use Solver to find the maximum likelihood estimates for the probability of detection, the probability of site occupancy, the probability of site extinction, the probability of site colonization, the probability of site suitability, the probability of a site becoming unsuitable, the probability of a site becoming suitable.
- To assess deviance of the saturated model.
- To compare model fit of various occupancy and suitability scenarios.
- To reinforce concepts of model fit.
- To learn how to simulate occupancy and habitat suitability data.

BASIC INFORMATION

The integrated model of habitat suitability and occupancy is an extension of the single–species, multi-season occupancy model introduced in Exercise 13. While the model is not yet available in Program PRESENCE or MARK, it will likely gain popularity as the need to explain observed occupancy estimates based on the suitability of habitat increases. This model was proposed in Chapter 10 of the book, Occupancy Modeling and Estimation (MacKenzie et al. 2006). We have developed this exercise with the purpose of demonstrating how integrated habitat suitability and occupancy modeling could be carried out in a multinomial maximum likelihood framework and to inform biologists of the type of field data needed to conduct this analysis.

Click on the worksheet labeled "Integrated habitat & occupancy" and let's get started.

BACKGROUND

This worksheet is set up similarly to the multi-season occupancy worksheet with an added level of excitement. Recall that the multi-season occupancy model allowed us to track changes in patch occupancy over time. The probability of a patch becoming extinct or colonized from season to season was incorporated to account for situations where either an occupied site became unoccupied (extinction) or an unoccupied site became occupied (colonization). Now consider a situation where, as a field biologist, you felt that, for your species of interest, year to year changes in site occupancy could possibly be explained by year to year changes in the suitability of that site. If the detection history was 10 00 11 00 10 across 5 years (or primary sampling periods), one scenario is that the species was present in year 1, in year 2 the species went extinct, the site was recolonized in year 3, went extinct again in year 4, and was recolonized in year 5. Although it is possible that the habitat could have remained suitable during years 2 and 4, you could imagine that it would be helpful to know if these extinctions were the result of the habitat becoming unsuitable. If the suitability of that site was characterized by year with a history of 1 0 1 0 1, where 1 = suitable, and 0 = unsuitable, it would be nice to apply a model that utilized BOTH histories (the species' encounter history as well as the site's suitability history) to determine if the species went extinct because the habitat was unsuitable.

To accomplish integration of occupancy and habitat, we will introduce a new state variable, $\xi(x_i)$, to represent the probability that a site x is suitable in sampling period i and η (eta) to represent a change in suitability. Additionally, we will modify the parameters used in the multi-season (Chapter 13) spreadsheet to account for a change in suitability state.

INTEGRATED MODELING OF HABITAT SUITABILITY AND OCCUPANCY

The integrated habitat suitability and occupancy model can easily become over complicated. To ensure that we get the concept across clearly, let's establish some additional assumptions of this model that can be relaxed, or reconsidered, later. All of the model assumptions mentioned in the single-species, multi-season occupancy model are the same here. As a reminder, these are: 1) The system is demographically closed to changes in the occupancy status of site during the secondary sampling period. At the species level, this means that a species cannot colonize/immigrate to a site, or go locally extinct/emigrate from that site during a primary period, but this assumption is relaxed between primary periods. 2) Species are not falsely detected. 3) Detection at a site is independent of detection at other sites.

For the integrated habitat suitability and occupancy model, an additional assumption is that there is no possible way a species can occur in unsuitable habitat. Thus, habitat suitability is defined as either suitable, 1, or not, 0. Therefore, when an occupied site becomes unsuitable, it must also go extinct. A second assumption for the integrated habitat suitability and occupancy model will force us to think back to the multi-season occupancy model terminology. In the multi-season model, we said that there were primary sampling periods (seasons) that were made up of secondary sampling sessions (surveys). The main idea was that the population was assumed to be closed within primary sampling periods and assumed to be open to extinction and colonization events between primary sampling sessions. Using this same logic, the second assumption for integrated habitat and occupancy modeling is that habitat suitability remains stable during primary sampling periods and that suitability can only change between them. So to recap, colonization, extinction, and suitability can change only between primary sampling periods; they cannot change between secondary sampling sessions.

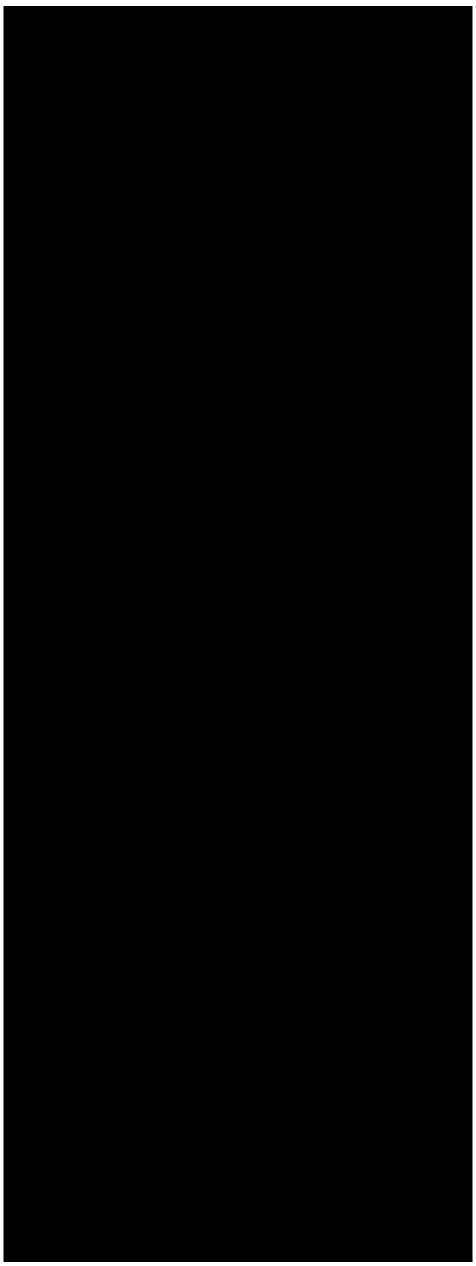
Now that we know the assumptions involved with defining habitat suitability, let's see walk through an actual example. The integrated model is useful for any situation where habitat does not remain suitable from season to season. An example that nicely demonstrates change in habitat suitability is the management of grassland bird habitat in agricultural regions. Available grassland habitat can undergo significant change on a year to year basis as crops are rotated, pastures are relocated, and, in a more finite sense, land is developed. For instance, a large field supporting several breeding pairs of Bobolinks one year may be plowed and planted with corn the next year. When you return to sample the site and find it has been planted with corn, not only will you record that no Bobolinks were detected, you must also note that the habitat was unsuitable. This example meets our assumptions because a corn field is unsuitable for Bobolinks and the change in suitability occurred between years and not within a year (where year is the primary period). Other examples of other situations where the integrated habitat and occupancy model may prove to be useful are: seasonal variation in vernal pool habitats, modeling populations within controlled wetland structures, or modeling populations sensitive to periodic drought.

SPREADSHEET INPUTS

[redacted]

Let's get oriented to the inputs, which are the encounter histories and their frequencies. The encounter histories are divided into primary and secondary sampling periods. Capture histories for the first primary period (year 1 for our grassland bird example) are given in cells C11:C35, and capture histories for the second primary period (year 2) are given in cells D11:D35. Some of the histories within a primary sampling period have brackets [ ] around them. The brackets indicate that the habitat was not suitable at that site for that primary period. You may also notice that all of the histories with brackets contain '00'. If you recall, one of the model assumptions mentioned above was that when a site is unsuitable, it cannot be occupied. Note that all combinations of the histories are given, of which there are 25. The frequencies of each history are given in cells E11:E35. The total number of sites surveyed is summed in cell E36, and the total number of unique histories is given in cell E37. The naïve occupancy estimate is the number of sites out of the total in which the species was detected.

THE SATURATED MODEL



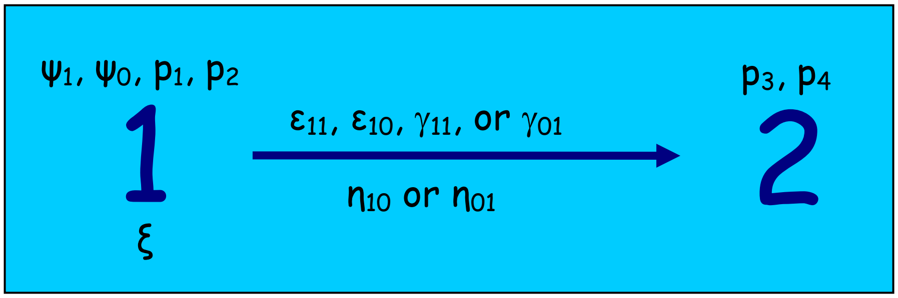
You might recall from both the single-season and multi-season occupancy models that we have all the information we need (the frequencies of each history) to compute the saturated model's -2Log_eL. Do you remember how we did it? The probability of getting a particular history is simply the frequency of that history divided by the total number of sites. The results give the MLE's for the raw data – they are simply the proportion of sites showing a particular history, and are computed in cells M11:M35. The sum of these cells must always be 1.0. The natural log of these probabilities is computed in cells N11:N35. To get the saturated model's Log_eL, we simply multiply the frequency of each history by the natural log of its probability, and then add the results up across histories. This result is multiplied by -2 to give the saturated model's -2Log$_e$L (cell N36), which you might remember is the basis for computing deviance. The formula in cell N36 is =-2*SUMPRODUCT (E11:E35,N11:N35). You can try plugging in different history probabilities to your heart's content, but you won't find a lower -2Log$_e$L than the one given in cell N36 because the probabilities fit the observed field data perfectly.

INTEGRATED MODELING OF HABITAT SUITABILITY AND OCCUPANCY PARAMETERS

This model has a lot of parameters to keep track of, so here is a cheat sheet for a model in which a species is surveyed over two primary seasons!

PARAMETER	DESCRIPTION
ψ_1	psi$_1$; the probability that a site is occupied at in season 1, given that it is suitable
ψ_0	psi$_0$; the probability that a site is occupied in season 1, given that it is unsuitable; this must always equal 0
p_i	The probability that the species of interest is detected at a site in survey i, given the site is occupied
ε_{11}	epsilon$_{11}$; the probability that a site will go extinct in season 2, given that it was suitable in both season 1 and 2
ε_{10}	epsilon$_{10}$; the probability that a site will go extinct in season 2, given that it is suitable in season 1 and unsuitable in season 2; this must always equal 1
γ_{11}	gamma$_{11}$; the probability that a site will become colonized in season 2, given that it is suitable in both season 1 and 2
γ_{01}	gamma$_{01}$; the probability that a site will become colonized in season 2, given that it is unsuitable in season 1 and then becomes suitable in season 2
ξ	xi; the probability that a site is suitable in season 1
η_{10}	eta$_{10}$; the probability that a site becomes unsuitable in season 2, given that it was suitable in season 1; (1-η_{10}) = the probability that a site will remain suitable in both season 1 and 2
η_{01}	eta$_{01}$; the probability that a site becomes suitable in season 2, given that it was unsuitable in season 1

To get a handle on where all of these parameters belong in the model and how they interact, we turn our attention to the blue diagram at the top of the spreadsheet.



The large 1 and 2 in the diagram indicate the first and second primary periods, which we can refer to as season 1 and season 2, respectively. Let's start with the season 1 and move our way to season 2. You should see that the parameters estimated in during the first season are ξ, ψ_1, ψ_0, p_1, and p_2...very similar to what we did in the robust occupancy model with the addition of ψ_0 and ξ. Here is where we estimate the probability of suitability (ξ) and occupancy (ψ_1 and ψ_0) and then whether the species of interest is detected or not (p_1 and p_2). Now notice the arrow pointing from primary period 1 to 2. All of the parameters associated with the arrow describe possible events that can occur between the first and second seasons. They represent changes in the state variables (occupancy and suitability). Let's begin by discussing the parameters that describe changes in occupancy (ε_{11}, ε_{10}, γ_{11}, and γ_{01}), keeping in mind that the 1s and 0s associated with each parameter are representative of the change in suitability state (between seasons 1 and 2), not occupancy. So, if a site is suitable and occupied during season 1, it can either go extinct with a probability of ε_{11} or it can remain occupied with a probability of (1-ε_{11}), given that the site remains suitable. Conversely, a suitable, occupied site during the season 1 can go extinct in season 2 as a result of that site becoming unsuitable with a probability of ε_{10} (always equal to 1 according to the assumptions of this model). If a site is suitable for both seasons but not occupied during the first, it can become recolonized in the second season with a probability of γ_{11}, or it can remain unoccupied with a probability of (1- γ_{11}). If a site is not occupied during the first season because it is not suitable, it can become occupied, if it becomes suitable in the second season with a probability of γ_{01}. Whether the species is detected or not in the second season depends on the site's suitability, occupancy state and the species detection probability in each survey (p_3 and p_4).

Now let's go over the parameters that describe changes in a site's suitability state between primary periods 1 and 2. If a site is suitable in season 1, it can become unsuitable with a probability of η_{10} or remain suitable with a probability of (1- η_{10}) in season 2. If a site is unsuitable in season 1, it can become suitable with a probability of η_{01} or remain unsuitable in season 2 with a probability of (1- η_{01}). Note that we don't directly estimate ψ_1, ψ_0 or ξ for primary period 2 because occupancy is determined by the state of the site in primary period 1, and whether the site went extinct, was colonized, remained occupied, or remained empty (just like the standard multi-season model). Similarly, suitability is determined during the first primary period (ε), and thereafter changes in suitability are determined by η. If a site was suitable in primary period 1, it can became unsuitable or remained suitable; if a site was unsuitable in primary period 1, it can become suitable or remained unsuitable.

Now let's focus on estimating the probability of each history, not from the raw data, but from the habitat suitability occupancy model parameters that we just discussed. The parameters for this model are listed in cells F11:F21.

	F	G	H	I
10	Parameter	Estimate?	Betas	MLE
11	p1	1		pe1
12	p2	1		pe2
13	p3	1		pe3
14	p4	1		pe4
15	ψ1	1		psi1
16	ε11	1		eps11
17	γ11	1		gam11
18	γ01	1		gam01
19	ξ	1		_xi
20	η10	1		eta10
21	η01	1		eta01

"Why isn't ψ_0 or ε_{10} included in the spreadsheet as estimated parameters?" you might ask. Since ψ_0 is the probability that a site is occupied given that the site is not suitable, it therefore does not need to be estimated. Similarly, ε_{10} is the probability that a site goes extinct due to its change from suitable in season 1 to unsuitable in season 2, and this probability must always equal 1.

In the above screenshot, the cell names for the MLE column are displayed. These cells are named pe1, pe2, pe3, pe4, psi1, eps11, gam11, gam01, _xi, eta10 and eta01. Click on one of the cells (I11:I21) and you should see the name of the cell appear to the left of the formula bar. These cell names are used in the spreadsheet for constructing the history probabilities and will be used in this section instead of the symbols used previously to minimize confusion. Additionally, in cells G11:G21 with the column heading 'Estimate?' a 1 or 0 is entered depending on whether a parameter is estimated or not. These cells are conditionally formatted and they will turn white when a 0 is entered. It is important to remember to fill this column in as it is used to compute K (the number of parameters estimated a given model; cell H30). Ultimately, our goal will be to find the combination of betas that maximizes the multinomial likelihood function. The model probabilities are linked to the betas once again with a logit link in cells I11:I21.

PROBABILITY OF EACH HISTORY

Now we're ready to write out the probability of obtaining a particular history (cells J11:J35). Let's start with the history 10 11 (cells C12:D12), which indicates that a species was detected in the first survey but not the second survey in season 1 and was detected in both surveys in season 2. It also indicates that the habitat remained suitable (notice no brackets) for the two primary periods. The probability of getting this history (cell J12) is =_xi*psi1*pe1*(1-pe2)*(1-eta10)*(1-eps11)*pe3*pe4. The site was suitable (_xi) and occupied (psi1) in the first primary period. It was detected in the first survey (pe1), and was not detected in the second survey (1-pe2). The habitat remained suitable (1-eta10) and the animal failed to go extinct between the first and second primary periods (1-eps11). It was then detected in the third survey (pe3) and again in the fourth survey (pe4).

How about the history [00] 01 (cells C25:D25)? The probability of getting this history (cell J25) is =(1-_xi)*eta01*gam01*(1-pe3)*pe4. The site was not suitable (1-_xi) in the first primary period, and therefore cannot be occupied so we do not need to include psi1, pe1, or pe2. The site then became suitable (eta01) and was colonized (gam01) but the species of interest was not detected during the third survey (1-pe3), but was detected during the fourth survey (pe4).

Let's try a trickier one, history 00 [00] (C34:D34). See if you can write out the probability equation in the space below:

The equation in cell J34 is =_xi*psi1*(1-pe1)*(1-pe2)*eta10+_xi*(1-psi1)*eta10. Here we have two options for this particular history. The first is that the site was suitable (_xi) in the first primary period and was occupied (psi1), but missed on survey 1 (1-pe1) and 2 (1-pe2), then habitat then became unsuitable (eta10) and therefore the site went extinct and was not occupied, OR (represented by the '+' sign) the habitat was suitable (_xi) in the first primary period, but unoccupied (1-psi1), and then the site became unsuitable (eta10) and could therefore not become recolonized. We don't need to address a change in occupancy here because it is assumed that if a site becomes unsuitable, it also goes extinct.

Now try history [00] [00]! (Check your answer with the formula in cell J35.)

This one is tricky but we encourage you to work your way through the remaining histories (cells J11:J35). After you've convinced yourself that the equations are correct, notice that the sum of the histories, calculated in cell J36 is 1.

THE MULTINOMIAL LOG LIKELIHOOD

Now it's time for everyone's favorite equation, the multinomial log likelihood. This equation is given in cell H24:

=E11*LN(J11)+E12*LN(J12)+E13*LN(J13)+E14*LN(J14)+E15*LN(J15)+E16*LN(J16)+E17*LN(J17)+E18*LN(J18)+E19*LN(J19)+E20*LN(J20)+E21*LN(J21)+E22*LN(J22)+E23*LN(J23)+E24*LN(J24)+E25*LN(J25)+E26*LN(J26)+E27*LN(J27)+E28*LN(J28)+E29*LN(J29)+E30*LN(J30)+E31*LN(J31)+E32*LN(J32)+E33*LN(J33)+E34*LN(J34)+E35*LN(J35). Click on this cell, then click somewhere in the formula bar and you should see the cells used in the equation "light up" - this often helps to see the pattern more clearly. This formula is exactly what we used in previous worksheets for the Log Likelihood calculations. This formula could have been substantially shorted by using the SUMPRODUCT function, but then you can't see the symmetry of the lit up equation.

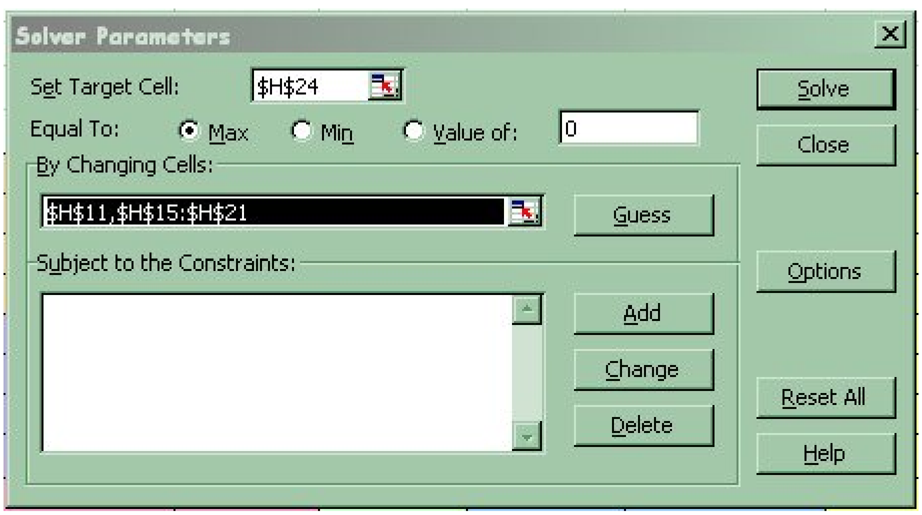
MAXIMIZING THE LOG LIKELIHOOD

The goal now is to maximize the multinomial log likelihood by changing the beta values in cells H11:H21 (and hence parameter MLE's). We are going to run the model $\psi_1\varepsilon_{11}\gamma_{11}(.)\gamma_{01}\xi\eta_{10}\eta_{01}p(.)$, in which we will estimate 8 parameters.

Don't forget to fill in a 0 or 1 in the 'Estimate?' column so K in cell H30 is computed correctly.

	F	G	H	I
10	Parameter	Estimate?	Betas	MLE
11	p1	1		0.50
12	p2	0	=H11	0.50
13	p3	0	=H11	0.50
14	p4	0	=H11	0.50
15	ψ1	1		0.50
16	ε11	1		0.50
17	γ11	1		0.50
18	γ01	1		0.50
19	ξ	1		0.50
20	η10	1		0.50
21	η01	1		0.50

Go to Tools | Solver, and set target cell H24 to a maximum by changing cells H11,H15:H21.



Press Solve and Solver will attempt to converge on an answer. Once Solver has found the betas, the MLE's that are linked to the betas are the maximum likelihood parameter estimates for p_1, p_2, p_3, p_4, ψ_1, ε_{11}, γ_{11}, γ_{01}, ξ, η_{10}, and η_{01}. Here are the results we got:

	F	G	H	I
10	Parameter	Estimate?	Betas	MLE
11	p1	1	0.251590867	0.62
12	p2	0	0.251590867	0.62
13	p3	0	0.251590867	0.62
14	p4	0	0.251590867	0.62
15	ψ1	1	0.289090582	0.64
16	ε11	1	0.025899206	0.51
17	γ11	1	0.645368487	0.80
18	γ01	1	0.374346036	0.68
19	ξ	1	0.355436744	0.67
20	η10	1	0.062354894	0.53
21	η01	1	0.279706425	0.64

Remember, when you run the model in MARK or PRESENCE, the programs report the estimate as well as the standard errors, and both are important. This spreadsheet does not include the standard errors for the various estimates. Let's look at these estimates more closely. In season 1, occupancy was 0.64, which is fairly high. Changes in occupancy are represented by ε_{11}, γ_{11}, and γ_{01}. The parameter ε_{11} is 0.51, indicating a 0.51 probability that a site will go extinct, given that the site is suitable in both time periods. Remember that ε_{10} (not shown) is always 1 because a site must go extinct if it becomes unsuitable. The probability that an empty site in year 1 will become colonized in year 2, given it is suitable in both years (γ_{11}) is 0.80. The probability that an empty site in year 1 will come colonized in year 2, given it was unsuitable in year 1 but suitable in year 2 (γ_{01}), is 0.68. In year 1, the probability of a site being suitable habitat was 0.67 (ε = 0.67). With regard to changes in suitability, η_{10} estimates the probability of a site going from suitable to unsuitable and is 0.53, while η_{01} estimates the probability of a site going from unsuitable to suitable and is 0.64. In other words, the study sites within this system are quite dynamic in terms of habitat suitability!

As this model is a spin off of the single-species multiple-season model (commonly used to model metapopulation dynamics), the following equation presented in exercise 13;

$$\psi_{t+1} = \psi_1 (1-\varepsilon_1) + (1-\psi_1) \gamma_1,$$

can be modified for the integrated habitat and occupancy model like so:

$$\psi1_{t+1} = \psi1_1 (1-\varepsilon11_1) + (1-\psi1_1) \gamma11_1 + (1-\psi1_1) \gamma01_1.$$

This equation allows us to calculate the probability that a site will be occupied in season t+1. It is the sum of sites that are currently occupied and do not go extinct ($\psi1_1$*(1-$\varepsilon11_1$)), unoccupied sites becoming recolonized in sites that remain suitable ((1-$\psi1_1$)*$\gamma11_1$) and that become suitable ((1-$\psi1_1$)*$\gamma01_1$). You may have noticed that the equation does not include the sites that go extinct because of the habitat becoming unsuitable ($\psi1_1$*(1-$\varepsilon10_1$)). This is because of the constraint of a site becoming extinct when it becomes unsuitable, causing ε10 = 1, making the term $\psi1_1$*(1-$\varepsilon10_1$) always equal to 0 and therefore unnecessary.

An estimate of the probability of suitability (ξ) of a given site at time t+1 can be derived from the change in suitability parameter in the following equation:

$$\xi_{t+1} = \xi_1(1-\eta10_1) + (1-\xi_1)\eta01_1$$

This is more similar to the equation presented in exercise 13 and mentioned above. The probability of a site being unsuitable in the 2nd primary period is the sum of the suitable sites that stay suitable ($\xi_1(1-\eta10_1)$) and the unsuitable sites that become suitable ($(1-\xi_1)\eta01_1$).

For the model that we just ran, $\psi1_2$ = 0.64(1-0.51) + (1-0.64)0.80 + (1-0.64)0.68 = 0.84 and ξ_2 = 0.67(1-0.53) + (1-0.67)0.64 = 0.52. You can see that these are both calculated below the model outputs on the bottom of the output on the spreadsheet.

MODEL OUTPUT

Now let's look at the remaining model output:

	G	H
23	**OUTPUTS**	
24	Log_eL =	-1411.58
25	-2Log_eL =	2823.16
26	Deviance =	14.577
27	Model DF =	17
28	C hat =	166.07
29	Effective N	500
30	K =	8
31	AIC =	2839.16
32	AICc =	2839.46

The model's Log_eL is computed in cell H24 (this is the cell you maximized), and the model's -2Log_eL is computed in cell H25. Deviance for this model is computed as the difference between this model's -2Log_eL and the saturated model's -2Log_eL with the equation =H25-N36. The model's degree of freedom is reported in cell H27 with the equation =E37-H30. It is simply the number of unique histories minus K (where K is the number of parameters estimated in cell H30). This computation goes back to the basic multinomial equation: we have 25 terms in the multinomial equation, and thus we estimate 25 probabilities in a saturated model. The model we just ran estimated those 25 probabilities indirectly by estimating 8 parameters: p(1=2=3=4), ψ1, ε11, γ11, γ01, ξ, η10, and η01. That leaves us with 25-8 = 17 parameters that are left over, or 17 degrees of freedom. AIC, AICc, and N are computed with the same equations as the general occupancy model.

ASSESSING FIT

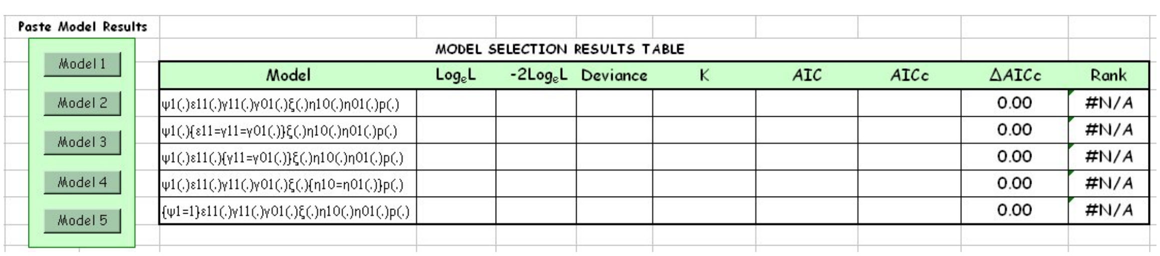
	K	L
9	Chi Square	
10	Expected	$(O-E)^2/E$
11	7.519	0.291515062
12	4.522	0.512175426
13	4.522	0.48320737
14	20.355	0.931931174
15	27.693	0.192144201
16	4.522	0.060220745
17	2.719	0.028994832
18	2.719	0.028994832
19	12.241	0.252833175
20	16.653	0.108890551
21	4.522	0.050564726
22	2.719	0.028994832
23	2.719	1.913056791
24	12.241	1.469201907
25	16.653	0.801468623
26	23.027	3.11058E-05
27	44.852	0.076456699
28	13.847	0.051830797
29	26.972	0.144137041
30	13.847	3.69482257
31	26.972	1.802061616
32	25.958	0.603579099
33	43.000	3.65555E-12
34	80.205	1.453000441
35	59.000	1.00764E-11
36	165.3	7.1542
37	p =	0.981462382

How well does this model fit? Well, we've entered formulae in cells K11:L35 to compute the Pearson Chi-Square. However, note that this approach has not been formally tested for the multi-season occupancy model, so it's still unclear whether this is a good measure of fit. Given this model's MLE's, the number of sites expected to have each encounter history is computed in cells K11:K35. This is exactly the same procedure we used to compute Chi-Square in the both the single and multiple season occupancy models: the probability of each history is multiplied by the total number of sites. The $(O-E)^2/E$ is computed in cells L11:L35. Remember, here is where you look for cell values greater than 3.8 – those with high values don't fit. The sum of $(O-E)^2/E$ is computed in cell L36 and is the model's Chi-

Square value. The p value associated with the Chi-Square value is computed in cell L37 with a CHIDIST function. This particular model appears to fit the data, as indicated by a p value much greater than 0.05. You could also run the MacKenzie and Bailey GOF but you'd have to modify your spreadsheet.

COMPARING MODELS

Since we have already compared the p(.) and p(t) models in previous exercises, here we are going to assume that p is constant for all surveys within the primary periods (i.e., p(.) model). Assuming this, we will compare models with varying combinations of the parameters of interest (ψ1, ε11, γ11, γ01, ξ, η10, η01). Of course, we will keep these scenarios biologically relevant to our grassland bird example. We just ran the model ψ1(.)ε11(.)γ11(.)γ01(.)ξ(.)η10(.)η01(.)p(.) and since we will be running a few more models, we want to save our results in the table labeled 'MODEL SELECTION RESULTS TABLE' so we can compare them after running the models in the following sections. Clear the model results table, except for the two columns that are responsible for ranking the models based on their AICc scores. Then, before moving on, click the button labeled 'Model 1' and you will see the results pasted in the model results table.



MODEL SELECTION RESULTS TABLE

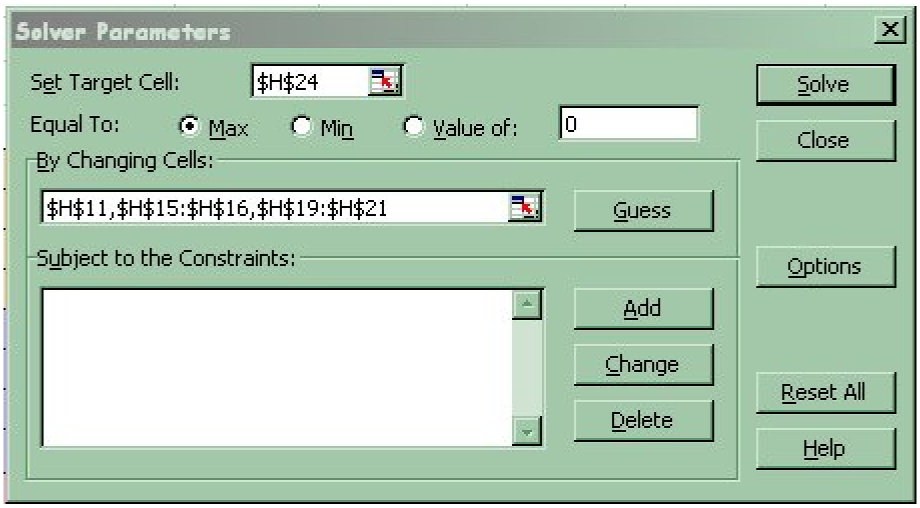
Model	Log_eL	$-2Log_eL$	Deviance	K	AIC	AICc	ΔAICc	Rank
ψ1(.)ε11(.)γ11(.)γ01(.)ξ(.)η10(.)η01(.)p(.)							0.00	#N/A
ψ1(.){ε11=γ11=γ01(.)}ξ(.)η10(.)η01(.)p(.)							0.00	#N/A
ψ1(.)ε11(.){γ11=γ01(.)}ξ(.)η10(.)η01(.)p(.)							0.00	#N/A
ψ1(.)ε11(.)γ11(.)γ01(.)ξ(.){η10=η01(.)}p(.)							0.00	#N/A
{ψ1=1}ε11(.)γ11(.)γ01(.)ξ(.)η10(.)η01(.)p(.)							0.00	#N/A

MODEL ψ1(.){ε11=γ11=γ01(.)}ξ(.)η10(.)η01(.)p(.)

This model forces the epsilon (ε) and gamma (γ) terms to be equal, signifying that recolonization and extinction are equal. Keep in mind that we are just setting three terms equal because ε10 is forced to be equal to 1 because when a site becomes unsuitable it will always be unoccupied. This can also be referred to as a metapopulation equilibrium model. For our Bobolinks, we test whether the rate of colonization and extinction are similar enough to assume a stable population that is neither declining nor increasing. In order to run the model, we need to force these three parameters to be equal to each other as well as entering a 0 in the 'Estimate?' column for the parameters that we will not be directly estimating. These cells are conditionally formatted to turn white when a 0 is entered into them.

	F	G	H
10	Parameter	Estimate?	Betas
11	p1	1	
12	p2	0	=H11
13	p3	0	=H11
14	p4	0	=H11
15	ψ1	1	
16	ε11	1	
17	γ11	0	=H16
18	γ01	0	=H16
19	ξ	1	
20	η10	1	
21	η01	1	

Next we will run solver, maximizing cell H24 by changing cells H11,H15:H16,H19:H21.



Press Solve and accept the Solver solutions. Paste the results to the model results table by clicking on the button labeled 'Model 2'. Once we have finished running all of the models we will analyze and compare the results more closely.

MODEL $\psi1(.)\varepsilon11(.)\{\gamma11=\gamma01(.)\}\xi(.)\eta10(.)\eta01(.)p(.)$

In this model, we will set the two recolonzation parameters equal to one another. With this model we are saying that a site will be recolonized, regardless of its prior suitability state (suitable or unsuitable). If suitable habitat is limited in our study system, the grasslands in an agricultural landscape, we would expect that if a site becomes suitable it will be occupied regardless of whether it was good habitat or not the season before. The set up is very similar to the previous model, except in this case we will just be setting the two gamma terms equal to one another as follows:

	F	G	H
10	Parameter	Estimate?	Betas
11	p1	1	
12	p2	0	=H11
13	p3	0	=H11
14	p4	0	=H11
15	ψ1	1	
16	ε11	1	
17	γ11	1	
18	γ01	0	=H17
19	ξ	1	
20	η10	1	
21	η01	1	

Now open up Solver and find the maximum value for cell H24 by changing cells H11,H15:H17,H19:H21. Press 'Solve', accept the Solver solutions, and then paste the results to the table by clicking on the button labeled 'Model 3'. The same type of model could be run for the change in suitability parameters, η10 and η01, by setting them equal to each other and testing to see whether a site is just as likely to become unsuitable as it is suitable. This is what we will do next.

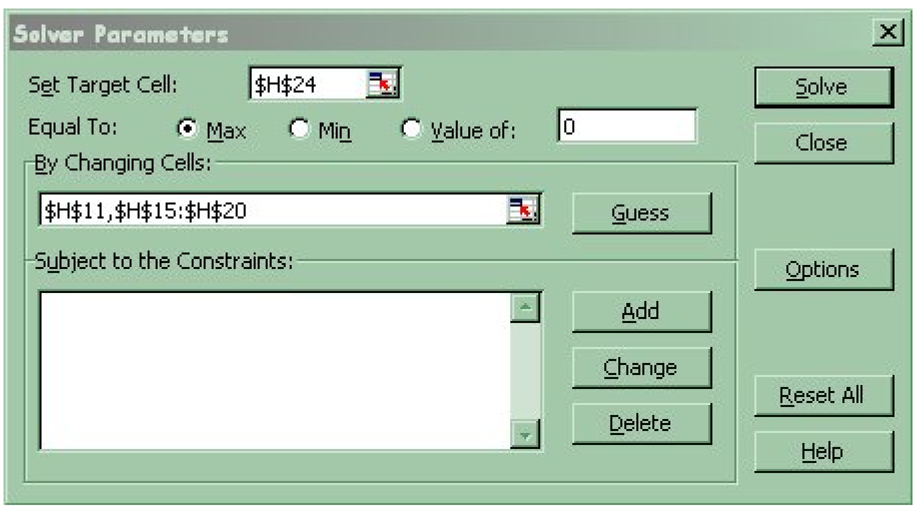
MODEL ψ1(.)ε11(.)γ11(.)γ01(.)ξ(.){η10=η01(.)}p(.)

Now, instead of adjusting the change in occupancy parameters, we will try a model in which we set the change in suitability state parameters equal to one another. This particular model is testing for support that a given site has an equal chance of becoming suitable as it does unsuitable. Keeping with our Bobolink example, this is basically saying that a corn field has the same chance of becoming a hayfield as a hayfield has of becoming a corn field.

Similar to what you did above, set η01 equal to η10 in the 'Beta' column (cell H21). Your set up should look like the one below.

	F	G	H
10	Parameter	Estimate?	Betas
11	p1	1	
12	p2	0	=H11
13	p3	0	=H11
14	p4	0	=H11
15	ψ1	1	
16	ε11	1	
17	γ11	1	
18	γ01	1	
19	ξ	1	
20	η10	1	
21	η01	0	=H20

Open Solver and set to maximize cell H24 by changing cells H11, H15:H20.



Press Solve and accept the Solver solutions. Then click on the button labeled 'Model 4' to paste the results to the model results table. After running the next and last model we will more closely examine the results table.

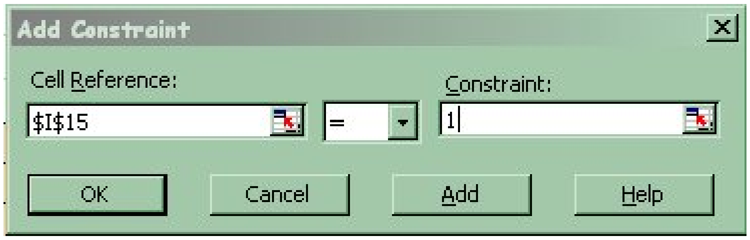
MODEL {ψ1=1}ε11(.)γ11(.)γ01(.)ξ(.)η10(.)η01(.)p(.)

Finally, we have come to the last model of the spreadsheet. In this model we force ψ1 to be equal to 1 by constraining it when we run Solver. This will test for support of the model that states: when a site becomes suitable, it automatically becomes occupied. In our study system, if this model were to receive substantial support, we would speculate that habitat is limited for grassland birds and therefore when a habitat patch opened up, it automatically becomes occupied. This model sets suitability and occupancy state equal to one another. Remember that ψ0 = 0, meaning that when a site is unsuitable it cannot be occupied, so in this model we are only looking at the occupancy status of suitable sites. To set up for running this model, place a 0 in the 'Estimate?' column for ψ1. We will be estimating all of the other parameters, except for p2-p4, which are set equal to p1 as in all of the other models we have run. Your spreadsheet should look like this:

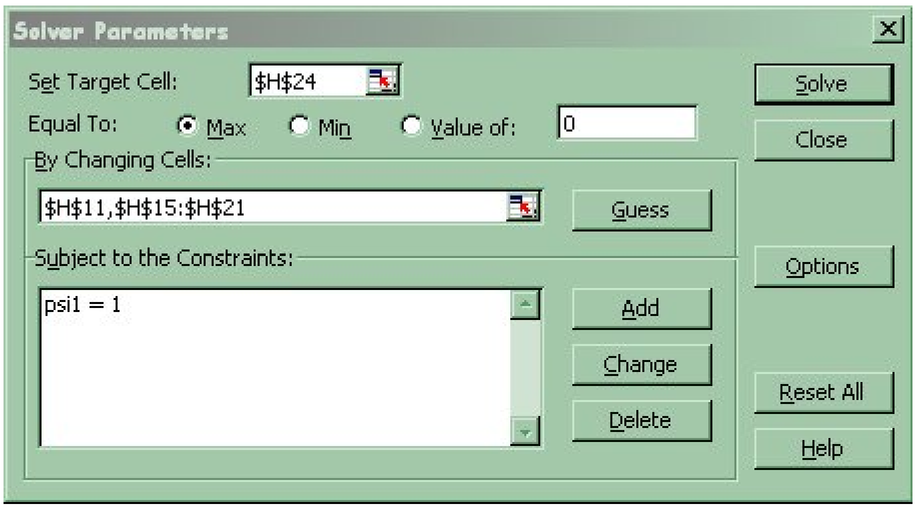
	F	G	H
10	Parameter	Estimate?	Betas
11	p1	1	
12	p2	0	=H11
13	p3	0	=H11
14	p4	0	=H11
15	ψ1	0	
16	ε11	1	
17	γ11	1	
18	γ01	1	
19	ξ	1	
20	η10	1	
21	η01	1	

Open Solver and find a maximum for cell H24 by changing cells H11,H15:H21. Here we also need to set ψ1=1 by adding a constraint. To do this, click on

'Add' next to the constraint box and another box will pop up, fill it in as follows:



This will force the MLE of ψ1 to be 1. Click OK. Your Solver set up should look like this:



Because the cell that contains the MLE is named, you will see psi1 instead of I15 in the constraint box. Click on Solve and accept the Solver solution. Paste the output to the model results table by clicking on the button labeled 'Model 5'.

Now it is time to look at all of the model output more closely.

	H	I	J	K	L	M	N	O	P
42	MODEL SELECTION RESULTS TABLE								
43	Model	Log_eL	$-2Log_eL$	Deviance	K	AIC	AICc	ΔAICc	Rank
44	ψ1(.)ε11(.)γ11(.)γ01(.)ξ(.)η10(.)η01(.)p(.)	-1411.58	2823.16	14.58	8	2839.16	2839.46	0.91	2
45	ψ1(.){ε11=γ11=γ01(.)}ξ(.)η10(.)η01(.)p(.)	-1415.90	2831.80	23.21	6	2843.80	2843.97	5.42	4
46	ψ1(.)ε11(.){γ11=γ01(.)}ξ(.)η10(.)η01(.)p(.)	-1412.16	2824.32	15.74	7	2838.32	2838.55	0.00	1
47	ψ1(.)ε11(.)γ11(.)γ01(.)ξ(.){η10=η01(.)}p(.)	-1414.16	2828.32	19.73	7	2842.32	2842.55	4.00	3
48	{ψ1=1}ε11(.)γ11(.)γ01(.)ξ(.)η10(.)η01(.)p(.)	-1450.09	2900.17	91.58	7	2914.17	2914.40	75.85	5

The models are ranked based on their AICc scores, so the model with the lowest AICc score has the highest rank among all of the models. In our table we see that the model that set the two recolonization terms equal is the most supported model given the data. This is what was expected as that is how we simulated the data, which will be the topic of the coming section. You may also notice that the AICc scores for the 1st and 2nd ranked models are very close (ΔAICc = 0.91), suggesting that there is support for both models, and the 3rd isn't very far behind (ΔAICc = 4.0). In this kind of situation, you would want to consider model averaging, which is a method for taking the parameter estimates from 1 or more substantially supported models by taking a weighted average. The weights are based on the support of each model included in the procedure. You can learn more about this techniques as well as more information about the model selection process in Model Selection and Multimodel Inference (Burnham and Anderson 2002).

SIMULATING DATA BY EXPECTATION AND WITH STOCHASTICITY

The last part of the spreadsheet covers how to simulate data for the integrated habitat and occupancy model. It is different from simulating robust occupancy model in that there are more parameters involved and we must incorporate the brackets into the histories. The constraints $\psi1=1$ and $\varepsilon10=0$ actually aid in simplifying the history probabilities by decreasing the number of possible scenarios.

	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
3	**Parameter Inputs**											
4	p1 =	p2 =	p3 =	p4 =	ψ1 =	ε11 =	γ11 =	γ01 =	ξ =	η10 =	η01 =	N =
5	0.6	0.6	0.6	0.6	0.7	0.5	0.8	0.8	0.7	0.5	0.6	500
6												
7	Individual	rand ξ	rand ψ1	rand p1	rand p2	rand η10	rand η01	rand ε11	rand γ11	rand γ01	rand p3	rand p4
8	1	0.251821	0.904349	0.442509	0.833464	0.982278	0.257386	0.358787	0.0411	0.4371	0.08446	0.78664
9	2	0.548033	0.149	0.342451	0.291516	0.884809	0.110648	0.890929	0.1896	0.2633	0.36066	0.65512
10	3	0.281771	0.326921	0.951324	0.558654	0.279905	0.834601	0.546415	0.4273	0.64	0.22431	0.95361
11	4	0.472956	0.530216	0.012082	0.445624	0.11011	0.205847	0.408786	0.4273	0.0632	0.83366	0.04495

First off, we enter the real parameter estimates that we want to use for simulating data in cells Q5:AA5 in addition to N, which is the number of sites you want to simulate data for. As in the other exercises, we will be simulating the data in two ways, with stochasticity and by expectation.

	AC	AD	AE	AF	AG	AH	AI
6	T1			T2			
7	ξa	p1	p2	ξb	p3	p4	History
8		1	0		0	0	10 00
9	[ ]	0	0	[ ]	0	0	[00] [00]
10		1	0	[ ]	0	0	10 [00]
11		0	1	[ ]	0	0	01 [00]

The first way involves a component of randomness to assigning encounter histories for each site. Then, for each site, we assign a random p1, p2, p3, p4, ψ1, ε11, γ11, γ01, ξ, η10, and η01 as shown in columns R:AB. The histories for each site will be based on these random numbers, and also the parameter estimates you specified previously. The habitat status of each site in the two primary periods is specified in columns AC and AF, respectively. The outcomes (0 or 1) of each of the four surveys (2 in each primary period) are listed in columns AD, AE, AG, and AH from 1 to 4, respectively. The site histories are generated in column AI by compiling all of the results from columns AC:AH. These are tallied up in cells AK8:32 and are what we used for the history frequencies to run the different models above.

First, the suitability of a site for primary session 1 is determined in column AC with the formula =IF(R8>Y5, "[ ]",""). This IF statement is saying that if the random ξ (column R) is larger than the specified ξ (cell Y5), then a set of brackets is entered into the cell (signifying that the site is unsuitable), otherwise it is left blank (representing a suitable site). Next the outcome from survey 1 (p1) in column AD is determined by the formulae =IF(AND(AC8="",S8<U5,T8<Q5),1,0). Breaking this IF statement with an imbedded AND statement down, we get: if the site is suitable (AC8="") AND it is occupied (S8<U5) determined by whether the random ψ1 is less than the specified ψ1, AND the species is detected (T8<Q5) determined by whether the random p1 is less than the specified p1. If all of these conditions are met, then a detection is recorded for that survey and a 1 is returned in the cell, otherwise, a 0 results signifying that the species was not detected, but does not mean that the site was not occupied. A similar formula is used to determine whether a detection results for survey 2 and is found in column AE, =IF(AND(AC8="",S8<U5,U8<R5),1,0).

Now things start to get a little more complicated. We are now at the point where both the change in suitability and the change in occupancy come into play. We will start with the suitability state in the 2nd primary period which is located in column AF and determined but the formula =IF(OR(AND(AC8="[ ]",W8>AA5),AND(AC8="",V8<Z5)),"[ ]",""). We can break it down by starting with the first term in the IF statement. Here we have 2 conditions that can be met so that would lead to the site being classified as unsuitable ([ ]), which is denoted by the 'OR'. The first set of conditions in the first AND statement say the site was unsuitable in the 1st primary period (AC8="[

]") AND then the site failed to become suitable (W8>AA5) determined by the random η01 being larger than the specified η01. The other combination of circumstances is that the site was suitable in the 1st primary period (AC8="") and that it became unsuitable in the 2nd primary period (V8<Z5) which is satisfied if the random η10 is less than the specified η10. If in either of the AND statements the conditions are met, the "[ ]" is returned and the site is unsuitable, otherwise the cell is left blank, indicating that it is suitable. Remember that the suitability of the site plays a role in determining occupancy, especially when the site is unsuitable.

Next, the outcome of the 3rd and 4th surveys are determined in columns AG and AH, respectively. Whether or not a site is occupied will depend both on the suitability state of the site as well as the change in occupancy parameters, ε11, γ11, and γ01. In column AG, the outcome of the 3rd survey is determined by the formula =IF(AF8="[ ]",0, IF(OR(AND(AND(AC8="", AF8=""), S8<U5, X8>V5,AA8<S5),AND(AND(AC8="[ ]", AF8=""), Z8<X5, AA8<S5), AND(AND(AC8="", AF8=""), S8>U5, Y8<W5, AA8<S5)), 1, 0)). Looking at this may make your head spin, so that is why we will break it down piece by piece. The first component of the first IF statement states that if the site is unsuitable (AF8="[ ]"), then a 0 is returned because the site cannot be occupied and if it is not occupied, the species cannot be falsely detected (one of the key model assumptions). If the site is suitable (column AF is blank), then we move to the next IF statement. Here we have 3 combinations of factors that could lead to a detection, and for ease of viewing, they have been color-coded. The first (in blue) is that the site was suitable in both primary periods (AND(AC8="", AF8="")) AND the site is occupied (S8<U5) or the random ψ1 is smaller than the specified ψ1 AND the site did not go extinct (X8>V5) or the random ε11 is greater that the specified ε11 AND the species was detected (AA8<S5) or the random p3 was smaller than the specified p3. The second possibility (in red) is that the site was unsuitable in the 1st primary period and then became suitable in the 2nd primary period (AND(AC8="[ ]", AF8="")) AND then the site was recolonized (Z8<X5) or the random γ01 is smaller than the specified γ01, AND the species was detected (AA8<S5) or the random p3 is smaller than the specified p3. The final combination of factors leading to a detection in the 3rd survey are noted by the green text. It states that the site was unsuitable in both primary periods (AND(AC8="", AF8="")) AND it was not occupied (S8>U5) determined by the random ψ1 being greater than the specified ψ1, AND the site becoming recolonized (Y8<W5) or the random γ11 being smaller than the specified γ11, AND that the species was detected (AA8<S5) or the random p3 is smaller than the specified p3. If any of the above combinations of conditions are fulfilled, then a 1 is returned and if not, a 0, indicating that the species was not detected, keeping in mind that does not equate to the site being unoccupied. Whew!

Now onto simulating data by expectation, which is simpler and more straight forward. There's nothing fancy about this at all: just enter the same old encounter history probabilities that we did earlier for each type of encounter history, and multiply by N. To make things easier, we copied the formula for each history probability and pasted it in blue type. These are the equations entered into cells AL8:AL32. Note the sum of the expected frequencies must equal N (whatever you specified in cell AB5).

	AL	AM
6	Expected	
7	Frequency	
8	7.9380	_xi*psi1*pe1*pe2*(1-eta10)*(1-elip11)*pe3*pe4
9	5.292	_xi*psi1*pe1*(1-pe2)*(1-eta10)*(1-elip11)*pe3*pe4
10	5.292	_xi*psi1*(1-pe1)*pe2*(1-eta10)*(1-elip11)*pe3*pe4
11	18.648	_xi*psi1*(1-pe1)*(1-pe2)*(1-eta10)*(1-elip11)*pe3*pe4+_xi*(1-psi1)*(1-eta10)*gam11*pe3*pe4
12	25.92	(1-_xi)*eta01*gam01*pe3*pe4
13	5.292	_xi*psi1*pe1*pe2*(1-eta10)*(1-elip11)*pe3*(1-pe4)
14	3.528	_xi*psi1*pe1*(1-pe2)*(1-eta10)*(1-elip11)*pe3*(1-pe4)
15	3.528	_xi*psi1*(1-pe1)*pe2*(1-eta10)*(1-elip11)*pe3*(1-pe4)
16	12.432	_xi*psi1*(1-pe1)*(1-pe2)*(1-eta10)*(1-elip11)*pe3*(1-pe4)+_xi*(1-psi1)*(1-eta10)*gam11*pe3*(1-pe4)
17	17.28	(1-_xi)*eta01*gam01*pe3*(1-pe4)
18	5.292	_xi*psi1*pe1*pe2*(1-eta10)*(1-elip11)*(1-pe3)*pe4
19	3.528	_xi*psi1*pe1*(1-pe2)*(1-eta10)*(1-elip11)*(1-pe3)*pe4
20	3.528	_xi*psi1*(1-pe1)*pe2*(1-eta10)*(1-elip11)*(1-pe3)*pe4
21	12.432	_xi*psi1*(1-pe1)*(1-pe2)*(1-eta10)*(1-elip11)*(1-pe3)*pe4+_xi*(1-psi1)*(1-eta10)*gam11*(1-pe3)*pe4
22	17.28	(1-_xi)*eta01*gam01*(1-pe3)*pe4
23	25.578	_xi*psi1*pe1*pe2*(1-eta10)*elip11+_xi*psi1*pe1*pe2*(1-eta10)*(1-elip11)*(1-pe3)*(1-pe4)
24	44.1	_xi*psi1*pe1*pe2*eta10
25	17.052	_xi*psi1*pe1*(1-pe2)*(1-eta10)*elip11+_xi*psi1*pe1*(1-pe2)*(1-eta10)*(1-elip11)*(1-pe3)*(1-pe4)
26	29.4	_xi*psi1*pe1*(1-pe2)*eta10
27	17.052	_xi*psi1*(1-pe1)*pe2*(1-eta10)*elip11+_xi*psi1*(1-pe1)*pe2*(1-eta10)*(1-elip11)*(1-pe3)*(1-pe4)
28	29.4	_xi*psi1*(1-pe1)*pe2*eta10
29	28.588	_xi*psi1*(1-pe1)*(1-pe2)*(1-eta10)*(1-elip11)*(1-pe3)*(1-pe4)+_xi*psi1*(1-pe1)*(1-pe2)*(1-eta10)*elip11+_
30	29.52	(1-_xi)*eta01*gam01*(1-pe3)*(1-pe4)+(1-_xi)*eta01*(1-gam01)
31	72.1	_xi*psi1*(1-pe1)*(1-pe2)*eta10+_xi*(1-psi1)*eta10
32	60	(1-_xi)*(1-eta01)
33	500.0000	

Why would you generate data based on expectation? Well, it happens to be a GREAT way to conduct a power analysis when you are in the design phase of your study. If you have some idea of what the parameter estimates might be, you can simulate data with varying sample sizes, run the analysis through MARK or PRESENCE (once these programs are designed to deal with integrated habitat and occupancy models, of course), and then critically evaluate how sample size effects the estimates in terms of bias (how far off where the estimates from MARK or PRESENCE compared to values entered in cells Q5:AA5) and precision (by examining the standard errors around each estimate). It's relatively easy to do, and is an important step in planning a study.

CONCLUSION

As you have worked through this exercise, hopefully the utility of integrated habitat and occupancy modeling has become apparent. The integrated model allowed us to explore the relationship between grassland birds and their dynamic habitat in an ecologically relevant manner. Not only were we able to account for variability in site occupancy, but we also saw how occupancy was related to habitat suitability. Good luck as you work to understand population dynamics in a changing landscape!