Lab 1 – Introduction to running survival models in Program MARK

The following are partial step-by-step instructions that correspond very closely to the materials presented in Chapters 3 and 4 of the Gentle Introduction. I've spelled everything out in great detail for most of Part 1. In Part 2, the basic steps for creating a project and building models will be very similar, so I've omitted a greater amount of detail so that you can learn through repetition.

Some logistics before we get started:

You should download the MARK data files from Blackboard or the PhiDot website and save them in an easy to navigate to place. I recommend bringing a thumb drive or external hard drive and working off this during class. In theory the files should stay on the hard drive of the lab machines and you could work off the same one every class, but this way you can take the files and your analyses with you and rest assured they won't be lost. The analyses we'll be running in lab shouldn't be so large that they'll run appreciably slower from a flash drive.

Part I – First Steps.

1. Open mark.
2. File -> New to start a new project.
3. From the Specification Window, make sure the "Live Captures" data type is clicked – this is what we will be using.
4. Give the project a title. This isn't terribly important now, but mark will tell us later that we should specify a title. Make it intuitive.
5. Next we need to import our input file. Click on the box to select a file next to the "Encounter Histories File Name" tab. Navigate to the Mark files folder and select the file "ED_MALES.INP" dataset, which contains capture histories for male European Dippers (*Cinclus cinclus*).
6. Click on the "view file" tab and Mark will open the text file containing the data. Notice the number of encounter occasions and the "staggered entry" of individuals. The "1" after the history tells us there was one individual with this capture history.

* Cue for a brief discussion of Input Files.

7. Next set the number of encounter occasions – note that you must do this as MARK does not "know" how to read that from the encounter history.
8. The second two tabs we will use in subsequent exercises where we incorporate group and covariate structures. For the time being leave "Attribute Groups" set to 1 and "Individual Covariates" set to 0.
9. When the pop up window tells you that ED_MALES.DBF has been created, click o.k. This is the file that will contain your analysis and all results.

* Cue to discuss dbase and naming models/input files.

10. The first thing you will see is a window with a triangular matrix of cells. This is the Parameter Index Matrix window for the survival parameters (Phi). You can open

additional Parameter Index Matrices by going to the PIM menu at the top of the page, and selecting the "Open Parameter Index Matrix" option. This will open a second window with a list of the other parameters, in this case the recapture parameter (p). Note that in this analysis we only have 2 parameters, but that will not always be the case.

11. In the PIM windows, each column represents an occasion (p) or an interval between occasions (phi). Each row represents a different cohort of individuals, beginning with those marked in occasion 1 (row 1), those marked in occasion 2 (row 2), etc. Notice there is a different value in each cell for each column, but not for each row. What these tell us is how p or phi is going to be modelled for each cohort in each occasion. Because the numbers vary among rows, we are saying that for each occasion (p) or interval (phi), the parameters are estimate independently. The values differ between the phi's (1-6) and p's (7-12) because they are distinct parameters that need to be estimated separately (i.e. we don't just count up starting at 1 within each parameter). The highest number in the list tells us something about the total number of total parameters we will be estimating – but in this case the total will actually be 11 because the last p and phi parameters are inherently confounded with each other.

* Cue to remind about confounding between p and phi in CJS models.

12. From the "Run" menu select the "current model" option. This will open a new window that allows you to set additional specifications for the model run. First, we need to specify a model name. Often students are confused here and think that we are telling Mark something about what we want it to run. That is not the case, as we already specified the model structure in the PIM windows. The name you give a model doesn't inherently matter (we could call it Cookie Monster Gumdrop; Mark doesn't care), but there is a standard convention that many chose to follow. This convention is highlighted in the 1992 paper by Lebraton et al. (cited in the class syllabus). For this present model, we would label it phi(t) p(t) where the "t" indicates that we've allowed survival and detection parameters full variation across time. Next, there is an option to specify the model link function. We'll discuss this more next week, for now leave the default of Sin link. The Var. Estimation box defines the method used to estimate the variance components of the models. The default is 2ndPart and what you'll typically want to use. The Fix Parameters box can be ignored for now, but it allows us to specify a specific value for certain parameters (e.g. p = 1.0) and there are sometimes logical reasons to do so. Those parameters are then not considered in the MLE, and their values may or may not influence estimation of other parameters. There is a list of other options that can be checked on the right, ignore them for now; we will cover some other relevant ones later.
13. Run the model by clicking the "OK to run" button. First, Mark will ask if we want to use the identity matrix. For now click yes, and next week we will be coving more advanced use of the design matrix where we won't rely on the default identity matrix.
14. A black dialog box will briefly open and then close – that was the numerical optimization happening and the likelihood being maximized. Don't blink, or you will miss it. The length of time needed to converge on a model depends on the complexity of the model. For most simple models it will chug along rapidly, but complex analyses from large datasets may take minutes, hours, or days to run.

15. Once the estimation runs, mark will close that dialog, open a window with some basic model results, and ask you if you would like to append the model results. Click yes.

* Brief pause to point out the warning message and discuss.

16. We now see that the model phi(t) p(t) has been added to a window called the results browser and we are given information on AIC, delta AIC, model weights, likelihood, parameters, etc. These are all results describing the fit of the model given our candidate set. Since we haven't run any other models, these results are largely meaningless at this point because all model selection statistics depend on comparisons with other competing models.
17. There are some additional results relevant to this first model that you can view at this point. Right click on the model name, and select the "Real Estimates" option from the dropdown tab. This will open a text file that displays the actual estimates for survival and detection probabilities for this model. From left to right the columns show the parameter number, the estimate, its standard error, and lower and upper bounds to the 95% confidence intervals. The parameter numbers correspond to the values you assigned to each parameter in the PIMs.

* Pause to review the information contained beginning on page 3 – 10 of the manual relating the PIM structure to the underlying parameter structures for the model

18. The first thing to look at in the model output here are the SEs and Confidence intervals. SE's that are unusually large, or CIs that go from 0 to 1, are suspect. In this case the SE's and CI's are having problems for the last phi and p estimates. Remember from last week and earlier today we discussed the inherent confounding between phi during the last interval and p in the final occasion (because there is no data from subsequent occasions to separately estimate them), so this isn't an inherently troubling issue. If you saw similar results for any other occasions, the implication is that your data are likely insufficient to estimate that parameter. Having parameters that will not estimate at this stage isn't the end of the world, but it does mean you will likely have to work with more constrained models and that could affect your ability to draw specific inferences from your data. We will discuss this in detail in the coming weeks.
19. Notice for the second p (detection probability during occasion 3) the estimate is 1.0, and the SE is extremely small with a CI that is basically 1.0. This has to do with that estimate being close to the upper bounds of the constrained value for the parameter. It is not an inherently bad thing, and the implication is that all (or almost all) individuals were detected during occasion 3. If you see something similar to this for a survival (or a similar biological parameter) you might be suspect. It means you may be estimating survival at an interval length that is not appropriate, or your sample was insufficient during that time period.

20. Ultimately we are interested in asking questions about survival, as much if not more so that simply generating estimates. We can do this by building alternative model structures by adjusting the PIMs, and interpreting the results using AIC. Model

structures that provide a stronger fit to the data will ultimately be supported in model selection, and will allow us make inferences about the processes that affect dipper survival. This will also have the side benefit of yielding parameter estimates that are closest to the true underlying values in the population. Based on what we've covered so far, go back and take a stab at running the following models on your own by changing the values in the cells of the PIM windows. Notice that the (.) notation signifies that a model is time-invariant, or constant.

Phi(t) p(.) Phi(.) p(t) Phi(.) p(.)

21. Realize that there are short cut operations for adjusting the PIM values. For example, you can right click and select the "constant" option to change all the values in the window to the same value as that contained in the upper left cell. Similarly selecting the "time" option will adjust the cells sequentially, again relative to the upper left cell. Run through these models on your own, and then we will look at the results and compare notes. Notice that there is an inherent level of parameter "accounting" you need to keep up with when adjusting PIMs. If you change the values in one window, you need to check to make sure you haven't inadvertently overlapped the indexing for another parameter type. If that happens you'll produce results that are nonsensical.

22. Finally let's run one additional model where we assume that during the 2nd and 3rd interval this system experienced unusually high flooding. We want to test the hypothesis that this flooding event negatively impacted dipper survival. To do so we need to change the indexing of the parameters to build a model where survival is estimated differently for intervals 2 and 3 compared with all other intervals. See if you can work out the appropriate way to change the PIMs to fit these last two structures on your own, and I can offer suggestions if you need help. Call this the phi (Flood) p (.) model.

23. After we look at the results of our 5 models we will also cover:

- Using the graphing feature in MARK
- The PIM Chart as an alternative for setting PIMS (have them replicate each of the 4 analyses using the PIM chart. (Point out the pitfalls of the "renumber with overlap button).
- General guidelines for starting new analyses and setting PIMs.

Part II – Fitting group models.

In this next exercise we are going to conduct a fairly similar analysis, but will use a new dataset collected from swifts in France that belong to two distinct breeding colonies. This group structure allows us to add an additional level of complexity in our analysis. Group structure is an important component of an analysis. It is the simplest way that you can include heterogeneity among individuals in your analysis, and it will form the basis for things like incorporating age effects later on.

1. Using the same procedures as outlined for the dipper data, create a new project from the encounter file aa.inp. Be sure to specify the correct number of occasions, and this time

specify 2 groups. You may elect to name them C1 and C2 or something similar. We are going to test the hypothesis that survival is inherently higher for the members of Colony 2.

2. Because we are now dealing with two potential sources of variance, group and time, we have a larger number of "simple" models that can be considered. These are depicted at the top of page 4-20. Notice that Cooch and White use the notation c*t to indicate models where colony and time are allowed to be interacting effects. Another conventional way to represent this would be as g*t, where the g stands for "Group". This is what I will use.
3. The default model is the fully general phi(g*t) p(g*t) model. Run this model for starters but don't worry about interpreting the results for now. Per instructions of Cooch and White, we will use the model phi(g*t) p(t) as our baseline "general" model. Using the PIM chart, create this model structure and run the model. Label it "phi(g*t) p(t) – CHART" or something similar.
4. Now we'll re-run the same model using the PIM Matrices instead of the chart, just to get some practice using both methods. First we need to re-set the PIMs to the most general model. In the results window, right click on the Phi(g*t) p(g*t) model and select the "Retrieve" option. A message should open telling you the model was retrieved successfully. Open the PIMs for each of the 4 parameters, which should now be set to be fully general. Adjust the PIMs to reflect the Phi(g*t) p(t) model and run it. Label it Phi(g*t) p(t) – PIM. You will know you did it correctly if the resulting model results are identical to those you got using the PIM Chart.
5. Using the method of your choosing, run the following additional models:

Phi(g*t) p(.) Phi(t) p(t) Phi(g) p(t) Phi(t) p(.) Phi(g) p(.) Phi(.) p(.)

Notice that this list is far from the total number of possible models we could run, but it will allow us to asses our hypothesis regarding colony-level variation in survival.

6. Once everyone is caught up to this point, we will go through the model selection results and the results for supported model structures.

PART III – Model Averaging.

We will finish the day with an example of generating model-averaged estimates parameter values, a concept we introduced last week and an operation that MARK performs quite conveniently for us.

1. Go to the "Output" option on the top of the page, select the "Model Averaging" open, and the "Real" option.
2. In the dialog box that opens, we are given the option to select from among all possible parameters by checking them off in a PIM-like matrix. Notice that we only really need to check the top row in this case – Why?
3. Select each of the relevant parameters for both groups and for both Survival and Recaptures. Be sure to click "estimates into excel" option, and leave any other boxes checked.

4. Mark will chug along and open two new windows. A text file describing the full model-averaging results, and an excel file with just the model-averaged estimates.
5. For comparison, let's pull the estimates from just the "best" model in our candidate set. Go to Output -> Specific Model Output -> Parameter Estimates -> Real Estimates -> and "Copy real estimates, SEs, and CIs into Excel".
6. We'll now work through the model-averaged and non-model-averaged results.

Lab 1 Assignment

On Blackboard under the Assignments Page you will find an input file labeled Lab1HW.inp. This contains a simulated capture history for two groups of hypothetical organisms. Assume these are small mammals and that the groups reflect two different color morphs, with group 1 consisting of "light" colored animals and group 2 consisting of "dark" colored animals. You are going to evaluate whether phenotypic variation in pelage color affects the annual survival of these fictitious rodents. Additionally, this study took place in a northern climate similar to that found in Maine. During the study there were two years of unusually mild winters, which occurred during intervals 2 and 5. You should evaluate whether survival changed during these years relative to other years with more "typical" winter weather. Finally, you should evaluate whether pelage color played a role in how sensitive each of the two color morphs of rodents were to changes in winter weather. That is to say, were light morph animals more or less affected by the mild winters compared to the dark morph animals? I will tell you that detection probability is constant through time, and differs among the two color morphs, so you can run all models assuming a p(group) structure. The suite of models you should run are:

1. Phi (Group*Time) p (Group)
2. Phi (Time) p (Group)
3. Phi (Group) p (Group)
4. Phi (.) p (Group)
5. Phi (Group*Winter Severity) p (Group)
6. Phi (Winter Severity) p (Group)

From your completed MARK analysis export your AIC table and your real parameter estimates, copy and paste them into a Word document, and upload this as your completed assignment. Also please attach the .dbf and .fpt files associated with your MARK analysis. Don't spend a lot of time worrying about formatting for you results tables, but please make sure what you submit is at least readable. The assignment is due prior to the start of lab next week.