

WILD 502 – Using Individual covariates in mark-recapture modeling

- Consider how you want to model each real parameter (φ & p), i.e., how will you constrain each as a function of covariates in the model.
- Set up the PIMs as you see fit after considering how many real parameters you want to separately model, i.e., how many rows will be in your Design Matrix and what parameters will they relate to.
- Consider whether you can (and want to) do the modeling by simply modifying the PIM values or whether you need (or want) to use the Design Matrix for modeling any of the real parameters. Remember: within a given model in your model-selection table, you are modeling both φ & p .
- Consider how many columns you'll need in the Design Matrix for your model structure. Remember that you'll need a set of columns that are uniquely used for φ & another set of columns that will uniquely be used for p .

Covariate types:

1. Time-varying covariate shared by all, e.g., snowpack or temperature
2. Group-level attribute – value shared by multiple animals, e.g., sex, age class
3. Individual covariate – can be a unique value or one shared by ≥ 1 animal, e.g., body mass

Can also consider whether covariates will be:

1. Categorical
2. Continuous
3. Time-varying or constant

We have already seen how we enter time-varying covariates shared by all: we set the relevant PIM up to be time-varying and enter the time-varying covariate values directly into the Design Matrix.

For a group-level attribute shared by multiple animals (but not all), we have to consider whether it is a continuous or categorical variable. If it's categorical, we can either put the values in the input file, or we can input encounter histories by group and inform MARK about the groups. If it's continuous, we'll often want to provide the values for the covariate in the input file and treat it as an individual covariate. The trade-off of using multiple groups has to do with ease of model building (you can build some models directly with PIMs and avoid using Design Matrix) versus having to deal with more PIMs.

Time-varying individual covariates can be very problematic in mark-recapture work because we often can't measure the covariate unless we handle the animal and we don't handle all animals on every occasion ($p < 1.0$). So, we end up with missing data. There are ways of dealing with this and you can read a bit about them in Chapter 11 of *Cooch & White*. But, if you're designing a mark-recapture modeling study, it's good to be aware of this issue and think through whether what you want is really feasible.

If you have a group-level attribute shared by multiple animals and choose to put the values in the input file, let's think about how you might do it. One easy way is with indicator variables that tells (indicates) whether or not the animal is in each of the various groups. You can assign group membership to n

groups with $n-1$ indicator variables, e.g., $\text{logit}(\varphi_{sex}) = \beta_0 + \beta_1 \cdot I(\text{Male})$, where $I(\text{Male})$ is an indicator variable that identifies whether an animal is male (yes=1) or not (no=0).

Model examples

Example 1

Apparent survival model: hypothesize that φ is positively related to water flow rate in the river

Detection rate model: hypothesize that p is positively related to trapping effort

The model has time variation in it, so we'll need both PIMS to be time-varying. The model does not have any individual covariates but will require us to enter the time-varying covariate shared by all into the Design Matrix. Note that the various beta's are **not** shared between different types of parameters. That is, in this example, B0 & B1 are only used to model φ & B2 and B3 are only used to model p

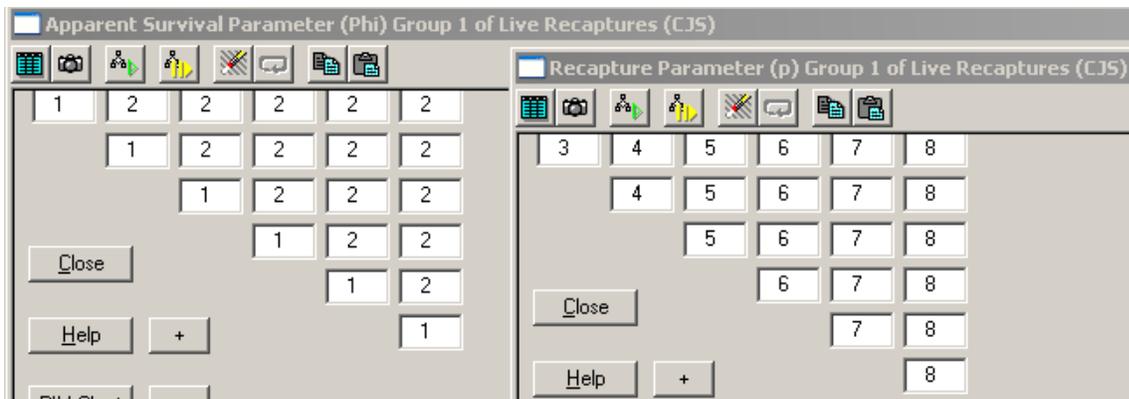
B0	B1	Param	B2	B3
1	12	1:Phi	0	0
1	9	2:Phi	0	0
1	4	3:Phi	0	0
1	15	4:Phi	0	0
1	11	5:Phi	0	0
1	8	6:Phi	0	0
0	0	7:p	1	2
0	0	8:p	1	3
0	0	9:p	1	2
0	0	10:p	1	4
0	0	11:p	1	1.5
0	0	12:p	1	5

Example 2

Apparent survival model: hypothesize that φ is positively related to fish body mass in 1st year of life and then was constant for older animals in all years. Imagine that all fish are caught as juveniles in summer.

Detection rate model: hypothesize that p is positively related to trapping effort.

The model for phi has no time-variation in it but does have age-related variation, so we'll need the PIM for φ to have 2 values related to age. The PIM for p will need to be time-varying and to not share any of the PIM values used in the PIM for φ . You might wonder if we need to concern ourselves with the possibility that p might be different for juveniles than older fish. You don't here because we don't estimate p for unmarked fish. So, fish initially marked as juveniles in year 1 will be adults in year 2 when you estimate a recapture rate (p) for them, i.e., you never estimate p for juvenile fish.



We would then need to call for a Design Matrix with 4 columns and to fill it in with appropriate values:

	B0 Phi-Intercept Young	B1 Phi-Mass Young	B2 Phi-Intercept Adults	Parm	B3 p-intercept	B4 p-effort
1		Mass	0	1:Phi	0	0
0		0	1	2:Phi	0	0
0		0	0	3:p	1	2
0		0	0	4:p	1	3
0		0	0	5:p	1	2
0		0	0	6:p	1	4
0		0	0	7:p	1	1.5
0		0	0	8:p	1	5

Example 3

Apparent survival model: hypothesize that ϕ is related to an individual's level of heterozygosity

Detection rate model: hypothesize that p is constant

Neither of the real parameters involves time-variation, so we can simplify the PIM values if we like. The Design Matrix will be quite simple.

B0	Parm	B1	B2
1	1:Phi	Heterozygosity	0
0	2:p	0	1

If you set each PIM to have constant values and to not share any values, you'll use a Design Matrix like this one.

B0	Parm	B1	B2
1	1:Phi	Heterozygosity	0
1	2:Phi	Heterozygosity	0
1	3:Phi	Heterozygosity	0
1	4:Phi	Heterozygosity	0
1	5:Phi	Heterozygosity	0
1	6:Phi	Heterozygosity	0
0	7:p	0	1
0	8:p	0	1
0	9:p	0	1
0	10:p	0	1
0	11:p	0	1
0	12:p	0	1

If, on the other hand, you leave the PIMs as time-varying, you'll need to repeat the model statements for each of the rows. But, note that this will provide the same results as the design matrix above because you're running the exact same model.

What's happening with the likelihood?

Remember back to our CJS probability statements for encounter histories when using time-varying parameters: $\Pr(111 \mid \text{release at period 1}) = \phi_1 \cdot p_2 \cdot \phi_2 \cdot p_3$

Well, here we're letting each of those parameters vary as functions of individual covariates. This is explained in detail on pages 430-431 of Williams et al. (section 17.1.7)

Let's consider our last model example:

$$\phi_j = \frac{\exp(\beta_0 + \beta_1 \cdot \text{Heterozygosity}_j)}{1 + \exp(\beta_0 + \beta_1 \cdot \text{Heterozygosity}_j)}; \quad p_i = \frac{\exp(\beta_2)}{1 + \exp(\beta_2)}$$

Now, each animal has its own apparent survival rate and shares a common recapture probability. So, working out the likelihood involves calculating the product of the probabilities of each individual's capture history over all individuals released during the study. One then finds (OK, one uses software to find), those values of β_0 , β_1 , & β_2 that maximize the likelihood and evaluates how certain of those values to be.

MARK provides you with several options for bringing in the covariate information: enter the actual covariate values (e.g., temperature) into the Design Matrix, provide covariate information by putting animals in different groups, provide covariate information for each animal by providing individual covariate values on the line of input where you enter that animal's encounter history information.

Chapter 11 of Cooch & White, especially the first 31 pages, provides a very good overview of this topic with special emphasis on how to do this in MARK. Chapter 17 of Williams et al. (2002) provides more details on the likelihoods, assumptions, and different parameterizations.

The power of these types of models to address biological hypotheses of interest makes them well worth the time it takes to learn them.

