CHAPTER 21

Density estimation...

Jake Ivan, *Colorado Parks and Wildlife*

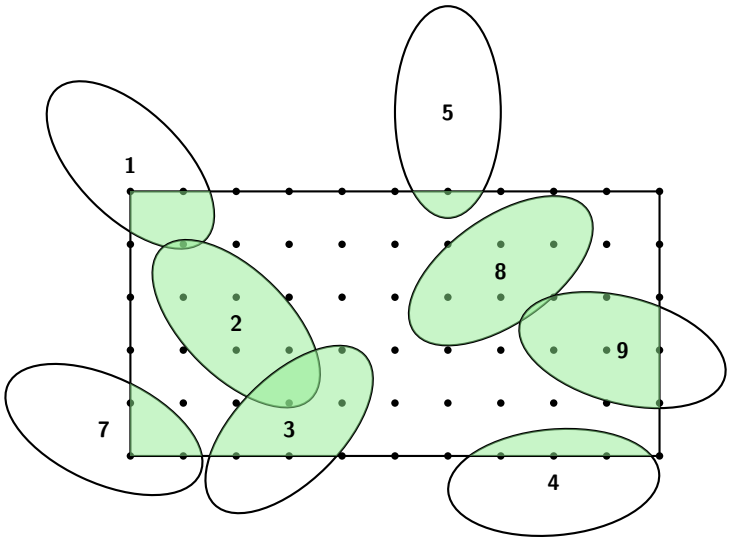
Abundance is commonly sought after as a state variable for the study of populations. However, density (number of animals per unit area) can be a more meaningful metric because it casts the state of a population into a common currency. For example, using closed capture models from Chapter 15, we may estimate 500 animals at site A and 200 animals at site B. Thus one conclusion we may reach is that habitat management at site A has positively impacted the population there compared to site B. However, if we know site A is 250 hectares and B is 100 hectares, then we realize that each has 2.0 animals/hectare. That is, on a relative scale, the different management scheme at A had no effect compared to site B. Conversely, we may estimate abundance at 2 sites to be similar and conclude management actions, or habitat types, or harvest regulations, etc. are having a similar impact, but if the sites are different sizes, then the impacts are actually quite different on a relative scale and our conclusion is erroneous. Thus, while abundance can be a useful metric, estimating density can be helpful as well.

Chapter 15 covered in depth how one might use mark-recapture techniques to obtain an estimate of abundance from closed population samples. And in this age of remote imagery, GPS, and GIS, it is relatively easy to delineate a study site and compute its area. Therefore, it should be easy to compute the number of animals per unit area, right? Simply obtain an estimate of abundance using any one (or several) of the multitude of models covered in Chapter 15, then divide that estimate by the area of the study site. Certainly you can proceed in this straightforward manner, and sometimes people do. However, your estimate will likely be biased. Why? Recall the assumptions of closed capture models. Not only do we need the marks to be individually identifiable, readable, and permanent (at least over the duration of the sampling), in order for our inference to be what we intend it to be, we need the population we're sampling to be closed both demographically and geographically.

Demographic closure can often be met by carefully considering the natural history of the species of interest, then timing the sampling to correspond to periods when births and immigration/emigration (e.g., dispersal) are unlikely. Sampling over a relatively short duration can give you some assurance that deaths during the sampling period should be relatively unlikely as well. However, achieving geographic closure is much more difficult. In fact, given that wildlife are mobile, we likely violate the geographic closure assumption in most applications. If you happen to be sampling a terrestrial species on a small island, you're probably fine. Or, inversely, if you're sampling fish in a pond, you probably meet the assumption in that case as well. You may even be able to reasonably assume closure in other select situations, such as sampling forest dwelling rodents in small woodland patches that are surrounded by a matrix of agriculture fields.

However, in the vast majority of situations, we sample wildlife populations in relatively continuous

habitat such that some animal home ranges (simplistically depicted below as ovals) are completely within the boundaries of our study site (i.e., the home range for animal 2 below falls completely within the study site [rectangle]), but many others only partially overlap it (animals 1, 3, 5, 6, 7, 8, 9). If this is the case, many individuals, especially those near the edges of the site, will move on and off of the area of interest *while mark-recapture sampling is ongoing*.



For example, during a traditional live-trapping study of small mammals in the fictitious situation depicted above, animal 9 may be on the site and available for capture during the 1st day of trapping (occasion 1), then off of the study area for day 2, back on for days 3 and 4, then off again on day 5. Assume for a moment that capture probability is perfect ($p = 1.0$). That is, whenever this animal is on the study area, it gets captured. Under this scenario, the encounter history for animal 9 would be '`10110`'. A closed capture model, however, assumes that if animal i was captured on the study area on any occasion, it was available on the study area for all occasions (i.e., the study area is closed – once an animal is on it, it can't move off). Given this assumption, the '`10110`' capture history would indicate that p is closer to 0.6 (this animal was available on 5 occasions and was detected on 3 of them). Thus, when we lack geographic closure, closed capture models will underestimate capture probability. In fact, the estimate returned from closed capture models is actually a product. It's the product of the probability that animal i is available for capture (a) and the probability that animal i is detected given it was available (p) such that $p_{\text{estimated}} = (p \times a)$. When we assume closure, we assume $a = 1.0$; when the assumption is violated $a < 1.0$ and $p_{\text{estimated}} < p_{\text{true}}$.

Now, recall from the beginning of Chapter 15 that the basic, heuristic estimator for abundance for a single occasion is:

$$\hat{N} = \frac{n}{\hat{p}},$$

where n is the number of unique animals observed (the count statistic) and $\hat{p}$ is the estimated probability that an individual is detected. It stands to reason that if $\hat{p}$ is underestimated when we lack closure, then our estimate of $\hat{N}$ will be inflated. Thus, when we fit a closed capture model to data from a design in which the study site is not closed, we do not obtain an estimate of the number of animals on the study site, which is what we want. Our estimate is larger than that. In fact, what we estimate is the number of animals that could have used the study site during the course of sampling. We term this the *super*

population (Schwarz & Arnason 1996, Kendall *et al.* 1997). Is this a problem? Maybe, maybe not. In some cases you may elect to simply re-define what it is you're estimating and report that number.

You can imagine, however, that if your goal is to estimate density, lack of geographic closure is problematic. We have an estimate of abundance, but we do not know area to which that estimate applies. Intuitively, the area used by the super population during the course of sampling was larger than the study area itself, but how much larger? In other words, in the expression for density

$$\hat{D} = \frac{\hat{N}}{A},$$

what value do we use for A? The choice is not clear, yet the choice can have a big effect on your estimate of density. Problem? Problem. What to do? Well, there are options. In fact, scientists have been proposing solutions to this very problem since the inception of the field of wildlife ecology in the 1930s. These include using home range estimates to take a stab at the effective area sampled (Dice 1938), exploiting differences in capture rates (or abundance estimates) between inner and outer detectors at the site (e.g., Maclulich 1951, Hansson 1969, Otis *et al.* 1978), using assessment lines to determine the reach of the initial sampling effort (e.g., Smith *et al.* 1971, Van Horne 1982), or computing the distances moved by individuals between detection events (Otis *et al.* 1978, Wilson and Anderson 1985). However, most of these approaches have fallen out of favor over time due to logistical issues, unrealistic data requirements, or *ad hoc* rather than theoretical foundations. A notable exception is the method based on mean maximum distance moved between trapping events (Otis *et al.* 1978, Wilson & Anderson 1985b), which has received criticism as an *ad hoc* approach (Williams *et al*. 2002, Royle & Dorazio 2008), and has shortcomings (Parmenter *et al.* 2003), but is still fairly popular (e.g., Converse *et al*. 2006, Zahratka & Shenk 2008).

Spatially explicit capture-recapture (SECR; Efford 2004, Borchers & Efford 2008, Royle & Dorazio 2008, Efford *et al.* 2009, Royle *et al.* 2009) and trapping webs (Anderson *et al*. 1983, Link & Barker 1994) are two contemporary density estimation approaches that circumvent the difficulties of estimating the area used by the super population by estimating density directly. These approaches generally have a stronger theoretical background behind them than the approaches mentioned above, but like any model, they have their own sets of assumptions and thus advantages and disadvantages. See Parmenter *et al*. (2003) and Ivan *et al*. (2013) for discussions of pros and cons, advantages, and limitations.

Here we focus on the solution provided to you in Program **MARK**. It is called the "**`Density with Telemetry`**" data type. From the name you can see right away that this approach requires auxiliary information (telemetry locations). Thus, as with the Barker model and Burnham's joint model, we are combining mark-recapture information with outside information to help solve a problem. The basic approach is exactly opposite of that taken during earlier work described above. That is, rather than estimating N, then trying to figure out the area to which the estimate applies, we first define the study site of interest, and thus fix its area. We then attempt to estimate the total number of whole and partial animals within its boundaries. In the case of our simple example above, this amounts to using telemetry to estimate the shaded areas for each of the 9 animals that were part of the mark-recapture data set. We then sum these proportions (i.e., for animals $1 \rightarrow 9$, we might estimate there are a total of $[0.47 + 1.00 + 0.08 + 0.00 + 0.37 + 0.93 + 0.09 + 0.82 + 0.07] = 3.83$ animals on the study site) and divide by the area of the study site to obtain an unbiased estimate of density.

The basic setup in the field is as follows. Assume that our sampling scenario is qualitatively similar to the simplistic example above. That is, despite our effort to nicely demarcate the study site of interest in a GIS, we know wildlife will go about their daily business largely ignoring our boundaries. Also assume we sample animals at the study site during a time of year in which we are not concerned about dispersal or births and we sample over a short enough time span that we're not so much concerned with deaths either (i.e., the population is demographically closed). During a sampling session, we capture,

mark, and release individuals on multiple occasions, let's say 5. Let's also suppose we fit these captured animals with a telemetry device (e.g., radio tag). We start collecting locations on telemetered individuals after we're done with the mark-recapture sampling and we sample locations from these animals over a relatively short period of time, say 2 weeks post-trapping. For each independent location, we note whether it is within the study site, or outside the study site.

We know we can apply traditional mark-recapture estimators to the mark-recapture data collected during the sampling session to estimate the super population (Schwarz & Arnason 1996, Kendall *et al.* 1997) of animals that used the site during sampling. Our goal is to use the telemetry data to estimate the portion of the super population that occurred within the boundaries of the site to produce an estimate of density corrected for the lack of geographic closure. Mathematically, we begin with the Huggins (1989,1991) closed capture estimator for abundance:

$$\hat{N}_{sp} = \sum_{i=1}^{M_{t+1}} \frac{1}{\hat{p}_i^*},$$

where $\hat{N}_{sp}$ is the abundance estimate that represents the *super population* of animals that could have used the site during the trapping session, $\hat{p}_i^*$ is the estimated probability that animal i is captured one or more times during the sampling session (i.e., if $\hat{p}_i$ is an estimate of the probability animal that i is detected on any given occasion, then $\hat{p}_i^* = 1 - (1 - \hat{p}_i)^t$, where t is the number of occasions), and M_{t+1} is the total number of animals detected. This should look very familiar if you've read Chapter 15, with the exception that we're explicitly noting that in most cases, $\hat{N}$ will be an estimate of the super population rather than an estimate of the number of animals within the study site.

Next let's make a substitution in the numerator of the Huggins estimator:

$$\hat{N}_{ss} = \sum_{i=1}^{M_{t+1}} \left(\frac{\hat{\tilde{p}}_i}{\hat{p}_i^*} \right),$$

where $\hat{\tilde{p}}$ is the estimated proportion of time (i.e., proportion of telemetry locations) animal i spent on the study site. Those individuals that are always located on the study site contribute fully to the estimate and are assigned $\hat{\tilde{p}}_i = 1$. That is, they are treated no differently than they are in the normal Huggins implementation. Most other individuals receive a fractional $\hat{\tilde{p}}_i$ because they spend some fraction of their time on the site. Depending on the circumstances, some (many?) animals may be like animal 4 above. That is, they may be attracted to our detectors and become part of the mark-recapture data set, especially if we use bait or lures to improve our detection probability, but in reality their normal home range doesn't occur within the study site. These animals will be assigned $\hat{\tilde{p}}_i = 0$ and do not contribute to the estimate.

So, functionally what we have here is an estimator that adjusts our count of animals detected upward (i.e., divide by $\hat{p}_i^*$) to accommodate imperfect detection, then ratchets that correction back down (i.e., multiply by $\hat{\tilde{p}}_i$) to include only those animals within the study area. $\hat{N}_{ss}$, then, is the estimated number of animals within the study site, which is what we've always wanted our abundance estimate to be.

At this point all we need to do is divide $\hat{N}_{ss}$ by the area of the study site (e.g., area of the minimum convex polygon encompassing all detectors. It is this polygon that is used to determine whether animals are 'in' or 'out', which informs derivation of $\tilde{p}$) in order to obtain an estimate of density:

$$\hat{D} = \frac{\sum_{i=1}^{M_{t+1}} \left(\frac{\hat{\tilde{p}}_i}{\hat{p}_i^*} \right)}{A},$$

where $\hat{D}$ is estimated density (number of animals per unit area), A is the area of the study site, and $\hat{\tilde{p}}_i$, $\hat{p}_i^*$, and M_{t+1} are as defined previously. **MARK** takes care of the variance calculation for you. Note that you do not necessarily have to put a telemetry device on each of the animals you sample via mark-recapture. If it is not possible to telemeter all individuals, we can use logistic regression to build a relationship between $\hat{\tilde{p}}_i$ and a suite of covariates for animals that do have a telemetry device. We can then use this logistic model to estimate $\hat{\tilde{p}}_i$ for those individuals that were not telemetered. See Ivan *et al*. (2013) for a summary of estimator performance for different levels of telemetry effort.

21.1. Likelihood

The likelihood for this estimator is simply a combination of likelihoods you have seen before. Chapter 15 has extensive detail regarding the structure of the likelihood for various forms of the Huggins estimator (i.e., the likelihood with respect to parameters p, c, and potentially π). Chapter 17 describes the likelihood for known fate survival models, which is a simple binomial – the same model used for logistic regression. In Chapter 17, the parameter we were trying to estimate was survival (S), success was defined as surviving an interval, failure was dying, and N, the number of trials (not to be confused with abundance, $\hat{N}$), was the number of animals alive to start each interval. In the case of density estimation, $\tilde{p}$ replaces S as the parameter of interest, a binomial "success" is locating an animal on the study site (i.e., the polygon), a "failure" is locating it off of the study site, and the number of trials, N, is the number of total locations collected (on an individual basis – the N's need not be the same for each individual). The two likelihoods can be written out separately, then multiplied together to form one big likelihood, which is then maximized.

21.2. Implementation in MARK

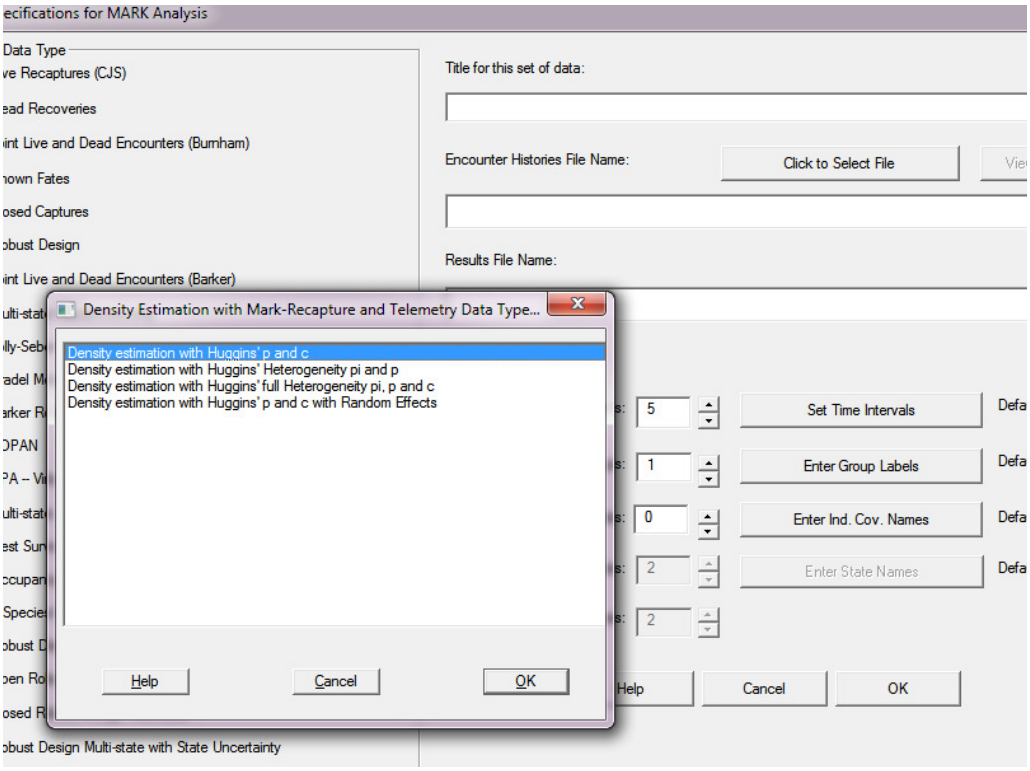
With this conceptual and mathematical background in mind, let's look at the mechanics of how to implement the model in **MARK**. First, the input file. As you may have guessed, it looks generally like a closed capture input file with a slight modification to accommodate the telemetry data. A piece of an example input file appears below.

As per the usual closed capture format, each line begins with an optional comment (here an animal ID #), followed by the encounter history (in this case, 5 occasions). Next are 2 new columns where we input the telemetry data. The first of these new columns is the number of times animal i was located on the study site; the second column is the total number of telemetry locations obtained on animal i. So, in the example, animal 20 was located on the study site 100% of the time (10 locations on site out of 10 total locations collected), animal 24 was located on the site 50% of the time (5/10), and animal 26 was never located within the bounds of the study site after having been captured and tagged there (0/10). Note also that animals 27 and 28 were part of the mark-recapture sample, but they were not fitted with telemetry devices. For those animals, we enter "." for each of the 2 telemetry columns as these animals provide no information for that part of the model. The remaining columns are the same as they would be for the usual closed capture input file. In this case, after the columns for the telemetry data, we have 2 groups (animals 20-24 belong to group 1, animals 25-28 belong to group 2), then a single covariate, which we will discuss later.

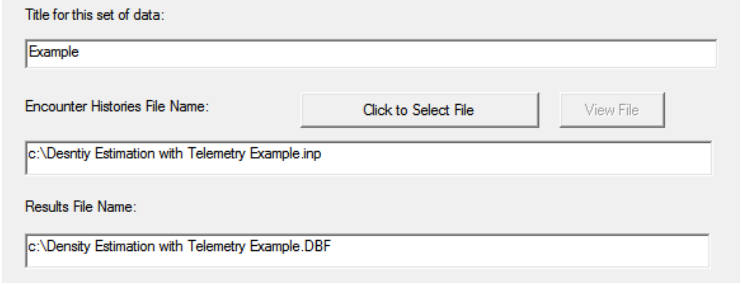
```
/*20*/        10010  10  10  1   0   20;
/*21*/        10000  1   10  1   0   0;
/*22*/        00100  8   10  1   0   20;
/*23*/        00001  10  10  1   0   20;
/*24*/        00010  5   10  1   0   0;
```

```
/*25*/        00010  5   10  0   1   20;
/*26*/        00010  0   10  0   1   0;
/*27*/        00110  .   .   0   1   25;
/*28*/        00001  .   .   0   1   0;
```

To load the input file, open **MARK**, choose "**`File | New`**" and select the "**`Density with Telemetry`**" data type near the bottom of the screen. A window appears in which you will have to select one of 3 possible parameterizations for the capture probability parameter(s). Note that these options are the same as those available to you under the traditional Huggins closed capture data type (Chapter 15). For the sake of this example, let's keep it relatively simple and choose "**`Huggins p and c`**":



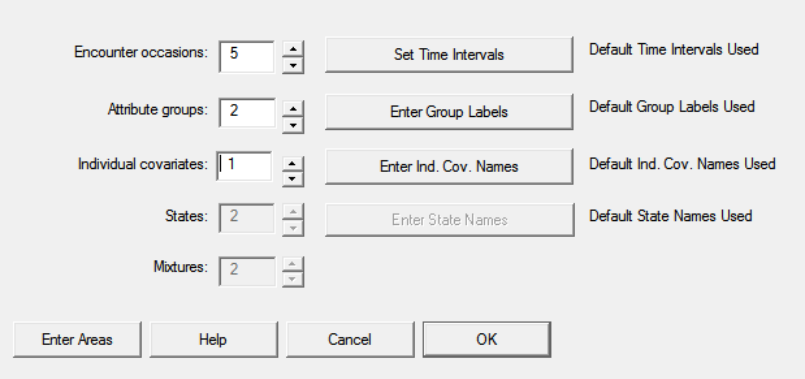
Once we've selected the parameterization we'd like to use to model capture probability, we need select an input file and fill in the rest of the specifications window as per usual. Let's title this data "`Example`", and choose the "**`Density Estimation With Telemetry.inp`**" input file.



This is simulated data intended to mimic a study of small mammals, such as deer mice, sampled in 2 sites (habitat types), A and B. Each habitat type was sampled with a (10×100) live-trapping grid (10m trap spacing). There are 5 occasions. In addition to marking each mouse with an individually identifiable ear tag, 50% of the individuals captured were fitted with a small VHF transmitter. These radio-tagged individuals were located once during the day and once at night for 5 days immediately after mark-recapture sampling ($n = 10$ locations total per animal) and each location was recorded as "in" or "out" of the study site.

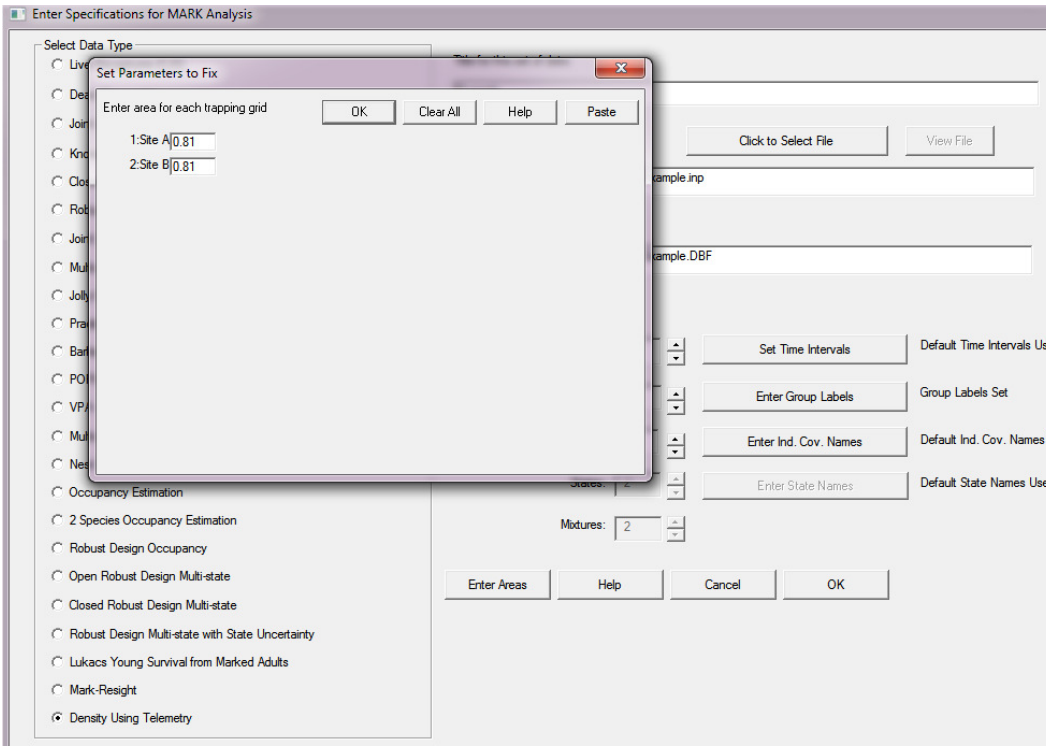
The single covariate we've recorded is the distance to the edge (DTE) of the of the site from the mean trap location of each individual (i.e., compute the mean trap location for each individual captured ≥ 1 time, then compute the minimum distance from this mean location to the edge of the site). This is a crude metric for characterizing whether an individual's home range is near the edge or in the interior of the site, which can be helpful in modeling both the p and $\tilde{p}$ parameters as we'll see below.

Given this information, fill in the remainder of the "**`Specifications Window`**" as follows:



Label the 2 groups "`Site A`" and "`Site B`"; label the Individual Covariate "`DTE`". Note that with this data type you are required to enter Group Labels. You cannot accept the defaults as you can with other data types. Also, note that there is a new "**`Enter Areas`**" button at the bottom of this window. You must click this button to specify the area of the study site for each group before you can proceed. The area of the site is critical to the final computation, and it defaults to 1. It is also allowed to vary with each group. Forcing you to double check the labels and corresponding areas before you proceed is **MARK**'s way of imparting some quality control on your project!

You should enter the area for each group in whatever units you would like the density estimate expressed. In our example (see top of the next page), each site (group) was sampled with a (10×10) grid with 10m trap spacing and we want the answer to be in animals per hectare so we enter 0.81 for Sites A and B (i.e., let's define the study site as the minimum convex polygon around the traps so the site is (90m$\times$90m). See - `sidebar` -, below, for more discussion.). If we wanted the answer to be animals/m^2, we would enter 8,100 for each group; for animals per kilometer, we would enter 0.0081 for each group, etc. If Site B was smaller, say 0.75 ha, we would enter 0.81 for the area of Site A and 0.75 for the area of Site B.



begin sidebar

what is a 'site'?

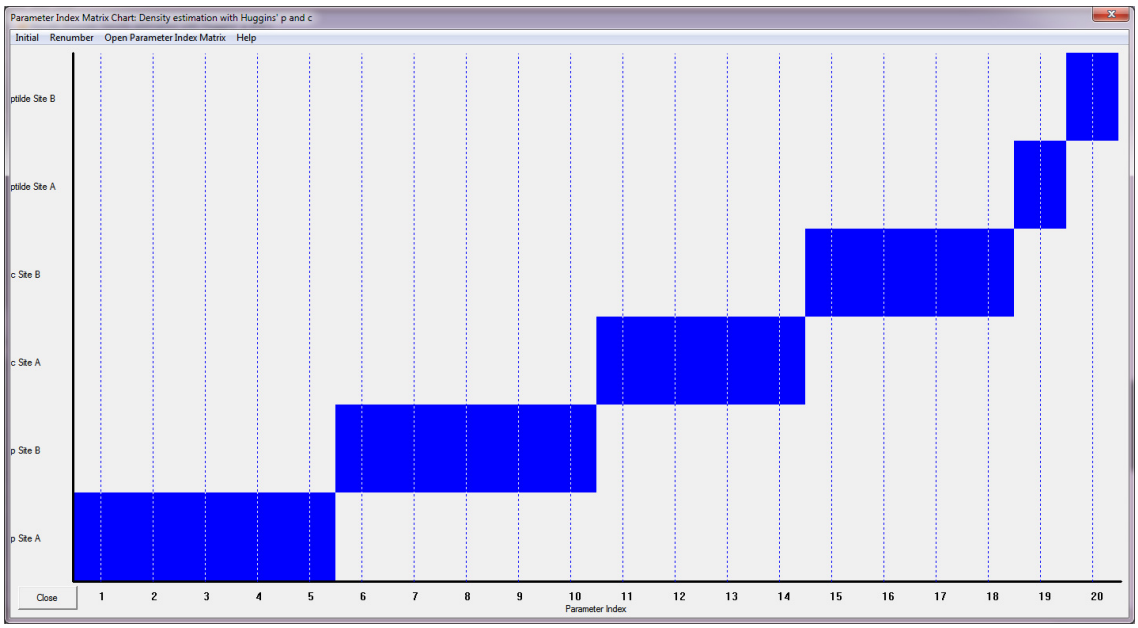
Note that the definition of the "site" is somewhat arbitrary. However, any workable definition should ensure that animals within the site have a reasonable chance of detection. We suggest that a "reasonable chance of detection" can be attained if the site is defined such that detectors are distributed at an adequate and roughly consistent rate (i.e., 4 detectors per home range; Otis *et al.* 1978, p. 76) within it so that there are no gaps in sampling effort.

We prefer to define the study site as the minimum convex polygon (MCP) encompassing the detectors. Such a definition seems natural and trap density inside this polygon would meet our criteria of being relatively high and consistent throughout. Alternatively, one could define the study site as the MCP plus ½ trap-width, or a full trap-width, and still claim that trap density is relatively high and consistent within the site.

However if we defined the site as the MCP plus 2-3 trap-widths or if detectors were distributed at a rate of 2 per home range over part of the site, but 6 per home range over other parts, those definitions would likely result in unequal sampling effort across the site and may lead to bias in the estimator. Note that because we incorporate the proportion of time each individual spends on the site (i.e., telemetry data), the estimator will be appropriate regardless of how the site is defined, as long as it is defined following the guidance we provide here.

end sidebar

Once you've entered the area for each group, **MARK** will allow you to click '**OK**' and proceed. In the main **MARK** interface, open the PIM chart (shown at the top of the next page) to view the parameters in the model. You will see the familiar p's (5 for each group) and c's (4 for each group) from the Huggins portion of the model along with the new parameter we're going to model using the telemetry data, $\tilde{p}$ (1 for each group).



Because the Huggins model is embedded within this model for density, all of the Huggins options from Chapter 15 are available to you for modeling p and c. That is, you can specify that capture probability is constant across animals and occasions (M_0), or that it varies by occasion (M_t), or that it differs depending on whether an animal has been captured previously (M_b), or that it differs generally among individuals (M_h), or that it varies in relation to one or more individual covariates. Of course, you can specify any combination of these basic model types as well. All of the nuances covered in Chapter 14 regarding model construction for closed abundance estimators (e.g., constraining the last p) apply here as well.

You may initially be tempted to structure similar models for $\tilde{p}$, but think before you start filling in your design matrix or PIM chart. For $\tilde{p}$ (proportion of time, or proportion of locations on the study site), there is no notion of a behavioral effect, nor time effects (we generally sample through time to estimate $\tilde{p}$, but $\tilde{p}$ is not connected to the mark-recapture sampling occasions in any way). Heterogeneity in $\tilde{p}$ is possible, and even likely, given that some animals will be located on the edge of the study site whereas others have home ranges near the interior. However, we cannot model general heterogeneity using a mixture model (π) like we do with capture probability. Instead, we're limited to the use of covariates, such as DTE. Thus, outside of covariates or groups, the structure we can apply to $\tilde{p}$ is limited.

Leave the PIMs in this general form, close the PIM chart, and open a '**reduced**' design matrix with 2 columns. Let's begin by building the simplest model possible: $\{p_. \tilde{p}_.\}$ (which we write in ASCII as `p(.)p~(.)`). For this first model, we're specifying a constant detection probability (M_0) and we're assigning each animal the average proportion of time on site across all animals and groups – akin to the basic form of this estimator first introduced by White and Shenk (2001). This may not be the most biologically realistic model we can think of, but it's a basis for comparison and may be the most supportable structure available if you have a sparse data set. The design matrix for this model is shown at the top of the next page:

B1: p int	Parm	B2: p~ int
1	1:p	0
1	2:p	0
1	3:p	0
1	4:p	0
1	5:p	0
1	6:p	0
1	7:p	0
1	8:p	0
1	9:p	0
1	10:p	0
1	11:c	0
1	12:c	0
1	13:c	0
1	14:c	0
1	15:c	0
1	16:c	0
1	17:c	0
1	18:c	0
0	19:ptilde	1
0	20:ptilde	1

Run this model then examine the real parameters. We see that capture probability was estimated to be 0.23 and $\tilde{p}$ was 0.66. So, on average, each animal was located on their respective study site about 2/3 of the time according to this model.

Example

Real Function Parameters of {p(.)p~(.)}

Parameter	Estimate	Standard Error	95% Confidence Interval Lower	Upper
1:p	0.2252994	0.0424134	0.1529914	0.3189148
2:p	0.2252994	0.0424134	0.1529914	0.3189148
3:p	0.2252994	0.0424134	0.1529914	0.3189148
4:p	0.2252994	0.0424134	0.1529914	0.3189148
5:p	0.2252994	0.0424134	0.1529914	0.3189148
6:p	0.2252994	0.0424134	0.1529914	0.3189148
7:p	0.2252994	0.0424134	0.1529914	0.3189148
8:p	0.2252994	0.0424134	0.1529914	0.3189148
9:p	0.2252994	0.0424134	0.1529914	0.3189148
10:p	0.2252994	0.0424134	0.1529914	0.3189148
11:c	0.2252994	0.0424134	0.1529914	0.3189148
12:c	0.2252994	0.0424134	0.1529914	0.3189148
13:c	0.2252994	0.0424134	0.1529914	0.3189148
14:c	0.2252994	0.0424134	0.1529914	0.3189148
15:c	0.2252994	0.0424134	0.1529914	0.3189148
16:c	0.2252994	0.0424134	0.1529914	0.3189148
17:c	0.2252994	0.0424134	0.1529914	0.3189148
18:c	0.2252994	0.0424134	0.1529914	0.3189148
19:ptilde	0.6647059	0.0362079	0.5904713	0.7316007
20:ptilde	0.6647059	0.0362079	0.5904713	0.7316007

The density estimate for the 2 sites is a derived parameter. If we examine the derived estimates (below)

Example 2

Estimates of Derived Parameters
Density Estimates of {p(.)p~(.)}

Group	D-hat	Standard Error	95% Confidence Interval Lower	Upper
1	21.626592	4.7100393	14.112441	33.141644
2	14.797142	3.7073721	9.0554150	24.179501

we notice that there appears to be some difference in density between the 2 study sites (21.63 animal/ha

in Site A, 14.80 animals/ha in Site B). Note, however, that because density is a derived parameter, we cannot specifically test for differences in density between sites by specifying a model structure that allows for differences and comparing it to one that does not. We can specify differences in p or $\tilde{p}$, but not $\hat{D}$. Instead we're left to make inference by other means. That is, we have to examine the point estimates, SEs, and confidence intervals, then make a judgment. In this case, the point estimates appear to be fairly different, but there is substantial overlap in the 95% confidence intervals, so maybe we conclude there is some evidence that density at Site A is higher than Site B, but the evidence is weak.

We might hypothesize differences in p and/or $\tilde{p}$ between the sites if the habitat imparts differential home range sizes and/or differential availability of food resources that impact detection. Let's go ahead and build 2 models that reflect these hypotheses, that p or $\tilde{p}$ may differ by site. Recall from above, that Site A (group 1) is represented by parameters 1-5 (p), 11-14 (c), and 19 ($\tilde{p}$); Site B (group 2) is represented by parameters 6-10 (p), 15-18 (c), and 20 ($\tilde{p}$). Thus, a potential model that reflects the hypothesis that p is different between sites might look like the design matrix on the left, whereas the one reflecting differences in $\tilde{p}$ might look like the one on the right. Of course, the 2 structures could be combined as well.

B1: p int	Parm	B2: p site	B3: ptilde int
1	1:p	0	0
1	2:p	0	0
1	3:p	0	0
1	4:p	0	0
1	5:p	0	0
1	6:p	1	0
1	7:p	1	0
1	8:p	1	0
1	9:p	1	0
1	10:p	1	0
1	11:c	0	0
1	12:c	0	0
1	13:c	0	0
1	14:c	0	0
1	15:c	1	0
1	16:c	1	0
1	17:c	1	0
1	18:c	1	0
0	19:ptilde	0	1
0	20:ptilde	0	1

B1: p int	Parm	B2: p~ int	B3: p~ site
1	1:p	0	0
1	2:p	0	0
1	3:p	0	0
1	4:p	0	0
1	5:p	0	0
1	6:p	0	0
1	7:p	0	0
1	8:p	0	0
1	9:p	0	0
1	10:p	0	0
1	11:c	0	0
1	12:c	0	0
1	13:c	0	0
1	14:c	0	0
1	15:c	0	0
1	16:c	0	0
1	17:c	0	0
1	18:c	0	0
0	19:ptilde	1	0
0	20:ptilde	1	1

In this case, the model that allows $\tilde{p}$ to vary between sites has more support than the one that allows p to vary between sites. If we examine the β estimates for the best fitting model thus far, we see that with Site A as the reference point (intercept), Site B tended to have lower $\tilde{p}$. That is, animals on that site tended to spend less time within the study boundaries compared to animals at the site represented by the intercept (Site A).

Example

LOGIT Link Function Parameters of {p(.)p~(site)}

Parameter	Beta	Standard Error	95% Confidence Interval Lower	Upper
1:p int	-1.2350466	0.2430015	-1.7113295	-0.7587636
2:p~ int	1.0459686	0.2279804	0.5991270	1.4928101
3:p~ site	-0.8163941	0.3314727	-1.4660806	-0.1667076

Finally, let's run a model that makes use of our DTE covariate. We might expect that if an animal is trapped, on average, near the center of the study site, it may have a higher probability of detection than animals trapped near the edge because there are more traps in its home range compared to the edge animal. Similarly, we might also expect such an individual to be located on the study site more often

than an animal whose home range is near the edge of the site. So, $\{p_{(DTE)}\,\tilde{p}_{(DTE)}\}$ (i.e., `p(DTE)p~(DTE)`) is a candidate model we may want to run, and we expect the DTE covariate to be positively related to both parameters. Since DTE is an individual covariate, we have to enter the covariate into the design matrix (see Chapter 11). For our present example, our DM should look like the following:

B1: p int	B2: p DTE	Parm	B3: p~ int	B4: p~ DTE
1	DTE	1:p	0	0
1	DTE	2:p	0	0
1	DTE	3:p	0	0
1	DTE	4:p	0	0
1	DTE	5:p	0	0
1	DTE	6:p	0	0
1	DTE	7:p	0	0
1	DTE	8:p	0	0
1	DTE	9:p	0	0
1	DTE	10:p	0	0
1	DTE	11:c	0	0
1	DTE	12:c	0	0
1	DTE	13:c	0	0
1	DTE	14:c	0	0
1	DTE	15:c	0	0
1	DTE	16:c	0	0
1	DTE	17:c	0	0
1	DTE	18:c	0	0
0	0	19:ptilde	1	DTE
0	0	20:ptilde	1	DTE

If we run this model, we see that our hypothesis was well supported. This model is just over 50 AIC_c units better than our previous best fitting model.

Model	AICc	Delta AICc	AICc Weight	Model Likelihood	No. Par.	Deviance
{p(DTE)p~(DTE)}	351.0997	0.0000	1.00000	1.0000	4	342.8416
{p(.)p~(site)}	401.1442	50.0445	0.00000	0.0000	3	394.9904
{p(.)p~(.)}	405.2051	54.1054	0.00000	0.0000	2	401.1286
{p(site)p~(.)}	405.5486	54.4489	0.00000	0.0000	3	399.3948

Also, as expected, the relationship between DTE and both p, and $\tilde{p}$ was positive. Note that this covariate can be computed for most any closed capture or density estimation setup. Keep it in mind as it often proves useful.

Example

LOGIT Link Function Parameters of {p(DTE)p~(DTE)}

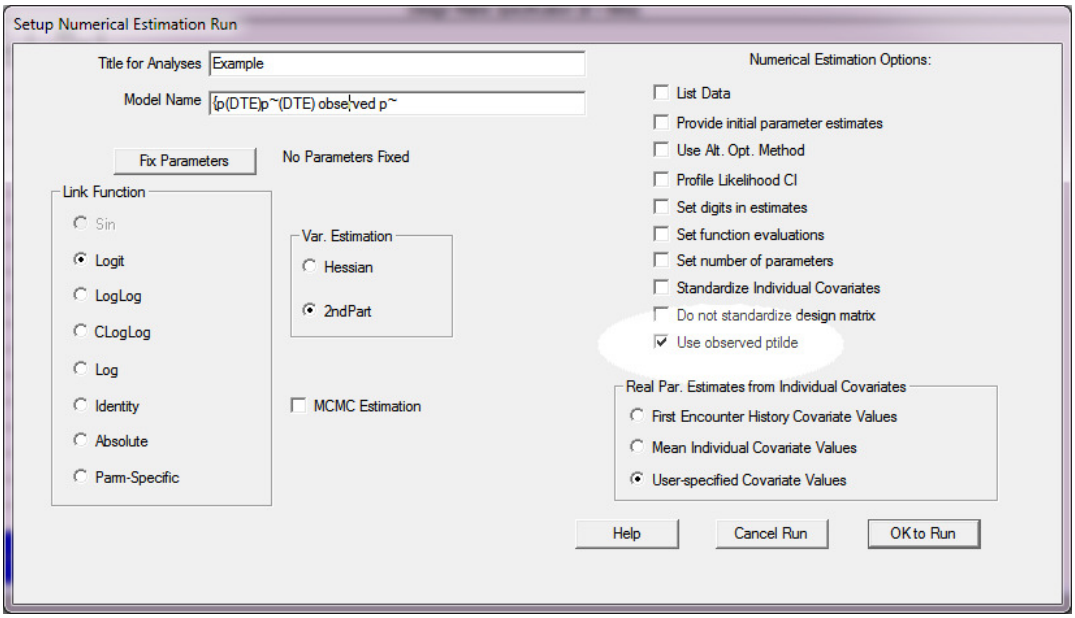
Parameter	Beta	Standard Error	95% Confidence Interval Lower	Upper
1:p int	-2.3480223	0.5668407	-3.4590301	-1.2370145
2:p DTE	0.0752421	0.0295541	0.0173160	0.1331682
3:p~ int	-0.6218991	0.2554951	-1.1226695	-0.1211287
4:p~ DTE	0.1450665	0.0241160	0.0977991	0.1923340

21.2.1. Estimate proportion on site or use the data?

To this point we've been accepting the default run options in **MARK**, such that $\tilde{p}$ is estimated using logistic regression. This means that we've been specifying a model that relates the proportion of time

on site to an intercept, or an intercept plus group indicators, or an intercept plus covariates, etc. **MARK** then uses this relationship to estimate $\tilde{p}$ for the animals that were sampled via mark-recapture but not telemetry (i.e., those individuals that had "`. .`" for the 2 telemetry columns in the input file) as *well* as those that were sampled via telemetry. In other words, the resulting logistic model gets applied to each of the i animals to derive an estimate of $\tilde{p}$ which is then summed over all individuals. You may be wondering why you wouldn't just use the telemetry data itself to estimate $\tilde{p}$ for those animals that were telemetered, then use a logistic model for only those animals that weren't telemetered.

Actually, you have that option. Retrieve model '`p(DTE)p~(DTE)`' in the browser and click the run button. Notice that in the run window (shown below) you have an extra option for this data type. You can check a box to "**`Use observed ptilde`**". Let's check the box this time and re-run the model. Add the label "**`observed ptilde`**" to the name so we can keep it straight.



Notice (shown below) that this model has the exact same number of parameters, same log likelihood, and same AIC_c as the original '`p(DTE)p~(DTE)`' model. This is because **MARK** still has to fit the logistic model to your data and used the MLEs to estimate $\tilde{p}$ for those animals that weren't telemetered. So, it still estimates the same number of parameters, it just doesn't use those parameters to estimate $\tilde{p}$ for some individuals. Thus, it doesn't make sense to compare the 2 approaches using AIC_c to see which is a more appropriate. It's completely up to the user and you should make a decision as to which way you'd like to go before you start running models, then run each model the same way.

Model	AICc	Delta AICc	AICc Weight	Model Likelihood	No. Par.	Deviance
{p(DTE)p~(DTE)}	351.0997	0.0000	0.50000	1.0000	4	342.8416
{p(DTE)p~(DTE) observed p~}	351.0997	0.0000	0.50000	1.0000	4	342.8416
{p(.)p~(site)}	401.1442	50.0445	0.00000	0.0000	3	394.9904
{p(.)p~(.)}	405.2051	54.1054	0.00000	0.0000	2	401.1286
{p(site)p~(.)}	405.5486	54.4489	0.00000	0.0000	3	399.3948

You might argue that since we're making inference using a model-based approach, we should be consistent and apply the model to all individuals. Or, you might argue that there is no sense in estimating

$\tilde{p}$ from a model when you can estimate it directly via sampling. Either approach is defensible. In practice, this decision often has little consequence (especially if nearly all animals are telemetered) and estimates are similar regardless of which flavor of model you prefer.

For example, if we compare the derived density estimates from the model we just ran to '`p(DTE)p~(DTE)`' we see that the estimates change slightly but we would probably make a similar inference regardless of the approach.

Example

Estimates of Derived Parameters
Density Estimates of {p(DTE)p~(DTE)} observed p~}

Group	D-hat	Standard Error	95% Confidence Interval Lower	Upper
1	22.042609	7.5809935	11.233331	43.253122
2	11.896178	4.8056765	5.3894304	26.258630

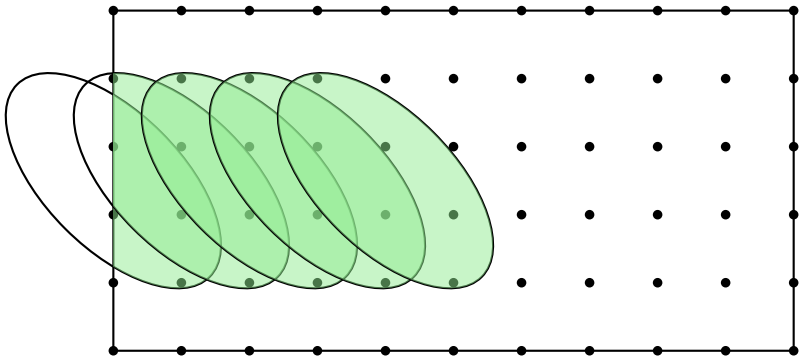
Example

Estimates of Derived Parameters
Density Estimates of {p(DTE)p~(DTE)}

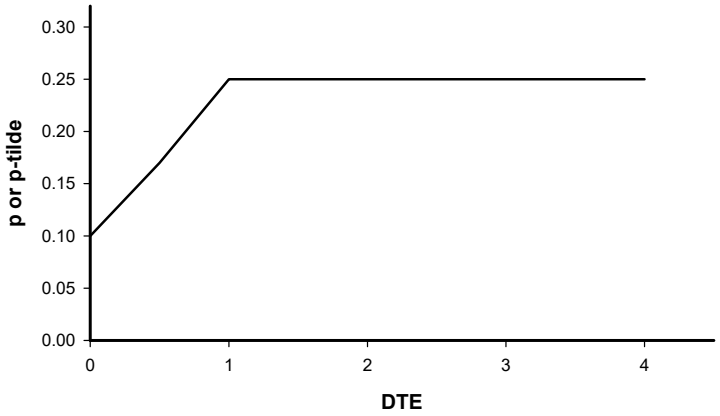
Group	D-hat	Standard Error	95% Confidence Interval Lower	Upper
1	20.476987	7.3514660	10.131404	41.386862
2	13.513596	4.9741462	6.5682701	27.802948

21.2.2. Threshold Model

Depending on the relationship between the size of your study site and the home range size of your target species, you may find that you only want the DTE covariate to operate up to a certain threshold. Consider the following example in which home ranges are now smaller than they were in our initial figure above.



As we move through the home ranges from left to right, once we encounter the 3rd or 4th home range, those animals are fully on the site and exposed to the same number of traps as those farther to the right (more toward the center of the site). In this situation, p and/or $\tilde{p}$ may increase with DTE for a given distance, but after that, DTE is no longer a very good predictor of either parameter. Thus, it may be useful to consider a "threshold model" for the estimation of $\tilde{p}$ and/or p such that the DTE covariate is only meaningful to a point. After that p and/or $\tilde{p}$ are constant.



Mathematically, such a model can be represented as

$$\text{logit}(\hat{\tilde{p}}_i) = \hat{\beta}_1 + \hat{\beta}_2(\min(\hat{\beta}_3, \text{DTE})),$$

where $\hat{\beta}_1$ and $\hat{\beta}_2$ are the usual intercept and slope estimates from a logistic model using distance to edge as a covariate, and $\hat{\beta}_3$ is the threshold parameter (i.e., the point at which the relationship between p and/or $\tilde{p}$ asymptotes). Note however, that you cannot build a model exactly like this in **MARK**. **MARK** is unable to estimate a parameter that is embedded within the '`min`' function. Instead, if you decide that a threshold model may be appropriate for your situation, build several models in which you give **MARK** an actual threshold value for DTE, then use AIC_c to help you figure out what the threshold should be.

In the example data set we've been analyzing, the trapping grid was (10×10) with 10m spacing. Thus, the center of the grid is 45m from any edge when the edge of the grid is considered to be exactly at the trap positions at the outer edge of the grid. Note that distance to the traps at the corners is in fact $(\sqrt{2} \times 45)$ from the center of the grid. Given that the edge of the grid is taken as the trap position on the outer edge, the threshold values we consider should be less than 45 (hopefully you can see that if we specify a model with a threshold of 45 in this specific example, we haven't specified a threshold at all – p or $\tilde{p}$ increase with DTE all the way to the center of the grid). Likewise, at the other end of the spectrum, we expect the threshold to be greater than at least the 1st trap width (remembering there should be 4 traps per home range). Thus, let's provide **MARK** with possible threshold values of 15, 25, and 35m for the p parameter. To do this, simply translate the equation above into a design matrix, making use of the design matrix function '**`min`**' (see Chapter 11).

Thus, the design matrix for a threshold of 15m looks something like:

B1: p int	B2: p DTE 15m	Parm	B3: p~ int	B4: p~ DTE
1	min(15,DTE)	1:p	0	0
1	min(15,DTE)	2:p	0	0
1	min(15,DTE)	3:p	0	0
1	min(15,DTE)	4:p	0	0
1	min(15,DTE)	5:p	0	0
1	min(15,DTE)	6:p	0	0
1	min(15,DTE)	7:p	0	0
1	min(15,DTE)	8:p	0	0
1	min(15,DTE)	9:p	0	0

For other thresholds, just replace the '15' in the design matrix expression with a different number (25 or 35). If we examine the results browser (below) we see that the 15m threshold is the best fitting model of the set we've constructed thus far:

Model	AICc	Delta AICc	AICc Weight	Model Likelihood	No. Par.	Deviance
{p(DTE 15m)p~(DTE)}	348.1651	0.0000	0.50762	1.0000	4	339.9071
{p(DTE 25m)p~(DTE)}	350.7230	2.5579	0.14129	0.2783	4	342.4649
{p(DTE)p~(DTE)}	351.0997	2.9346	0.11703	0.2305	4	342.8416
{p(DTE)p~(DTE)} observed p~}	351.0997	2.9346	0.11703	0.2305	4	342.8416
{p(DTE 35m)p~(DTE)}	351.0997	2.9346	0.11703	0.2305	4	342.8416
{p(.)p~(site)}	401.1442	52.9791	0.00000	0.0000	3	394.9904
{p(.)p~(.)}	405.2051	57.0400	0.00000	0.0000	2	401.1286
{p(site)p~(.)}	405.5486	57.3835	0.00000	0.0000	3	399.3948

21.3. Confidence Intervals

Recall from Chapter 15 that **MARK** computes log-based confidence intervals for abundance, partly due to lack of asymptotic normality in point estimates and partly due to the idea that we want the lower bound to be no less than M_{t+1} (makes no sense to have a lower confidence limit that is smaller than the number of animals you know are out there!). Superficially, we'd like to invoke a similar approach with density, and for the same reasons. Notice in section 14.1.10 that M_{t+1} (the number of animals capture or the minimum bound on the abundance estimate) appears in the calculations as does f_0, which is simply $\hat{N} - M_{t+1}$. You may be tempted to conclude that the analog to M_{t+1} for density estimation would the sum of all the $\tilde{p}_i$ divided by the area of the study site. Then the analog to f_0 would be $\hat{D} - (\sum \tilde{p}_i)/A$ and all of the same computations would follow. However, $\sum \tilde{p}_i$ is not a known quantity as M_{t+1} was in the abundance world. It's just an estimate (i.e., it's really $\sum \hat{\tilde{p}}_i$).

Thus, instead of following the closed capture CI computation exactly, **MARK** uses log-based confidence intervals that ensure the CIs are > 0, which is the minimum condition that has to be true. As such the 95% CIs are computed as:

$$\pm \exp\left[\log(\hat{D}) \pm 1.96\left(\frac{SE(\hat{D})}{\hat{D}}\right)\right].$$

21.4. Assumptions

Recall from Chapter 15 (and earlier in this chapter) that closed capture models require 2 assumptions: (1) the population is closed geographically (animals do not move on and off of the study site) and (2) demographically (there are no births, deaths, immigration or emigration). Through the use of auxiliary telemetry data we are able to relax the first of these assumptions when estimating density, but we still need the population to be closed demographically. Fortunately, demographic closure is usually much easier to attain and simply requires sampling for a relatively short duration at a time when you don't expect births or dispersal events. However, the density estimation model described here requires 3 additional assumptions beyond the usual closed capture ones:

1. The animals sampled with telemetry are representative of the population of animals that use the study site.

2. Telemetry devices do not affect movements and there are no effects of mark-recapture sampling on animal movements.
3. Error in telemetry locations is small relative to the size of the study site and assignment (on/off) of locations near the edge of the site is unbiased.

The crux of assumption 1) is that you want to avoid oversampling "interior" animals while under-sampling "edge" animals or vice-versa. Hopefully if you think back the form of the estimator, you can see why either of these cases would be problematic. If your severely under-sample edge animals (which should have relatively small values for $\tilde{p}_i$), the summation of $\tilde{p}_i$ in the numerator of our density expression will be too large (all we sampled were animals with $\tilde{p}_i$ near 1.0), and you will end up with an estimate of density that is biased high to some degree. The reverse could also be true, and would result in negatively biased results, although it is probably more likely that you would capture too many interior animals compared to too many edge animals.

Note, however, we've also previously determined that DTE is often a good predictor of capture probability. That is, interior animals may well have higher capture probabilities (p) than edge animals because they tend to have more traps or detectors in their home range. If this relationship holds in your study then it may be true that you tend to sample interior animals too often, which inflates the numerator of our density expression, but we're probably inflating the denominator at the same time, thereby canceling at least some of the potential bias. Nevertheless, you want to provide yourself with as much protection from bias as possible, and there are a couple of design features of mark-recapture sampling that may help.*

For assumption 2, the main concern is if baits or lures were used to detect animals during the mark-recapture portion of your study, which is commonplace. Such attractants are designed to alter the normal movements of animals (hopefully they entice animals to your detectors and increase your capture probability!), but when we're sampling animal movements using telemetry we'd like animal movements to be "natural". We have several recommendations for telemetry sampling when baits or lures were used during mark-recapture.

First, we recommend discarding from the analysis any telemetry locations obtained during the mark-recapture session as it is very probable that baited detectors influenced animal movement. Second, it is imperative that you the investigator ensure that every last morsel of bait has been removed from the site at the end of the mark-recapture session so there is no unnatural attractant left to influence movements after the session. Third, we suggest that it may be appropriate to wait 1–2 days post mark-recapture before collecting location data to allow animals to revert to their normal activity patterns. Telemetry sampling should, however, be completed within a reasonable time to avoid biasing estimates of $\tilde{p}_i$ due to seasonal movements, migration, or dispersal.

In summary, our experience suggests that bait and lures can have quite an impact on the movements of animals during a mark-recapture study, even enticing animals to make extensive movements to become part of the mark-recapture data set when their usual home range does not include the study site at all (e.g., animal 4 in the original example). Following the guidelines outlined here, telemetry information should result in $\tilde{p}_i = 0$ for these individuals, and they will not contribute to the density estimate, which is appropriate.

* For traditional live-trapping studies in which telemetry devices are deployed on a subset of animals during the mark-recapture sampling, we recommend checking traps in the same order on each occasion, but selecting a different random starting point each time. This strategy should help equalize the probability of capturing and telemetering edge vs. interior animals. In addition, we suggest retaining some telemetry devices for deployment during the latter portion of a sampling session to facilitate the inclusion of trap-shy individuals in the radio-tagged sample in addition to trap-happy individuals that are captured early and often.

All of this 'hand waving' is also a justification for using the logistic regression model for $\tilde{p}$ rather than the observed values, because the logistic regression model would correct for the bias of a non-random sample of animals that were radioed, assuming that at least a few edge animals (if not a random sample) made it into the telemetry sample.

21.5. Summary

The summary for Chapter 15 states that "Despite a seemingly simple goal, estimating abundance can be quite difficult...". The same can be said for density estimation due to all of the numerous, subtle complications embodied in the closed capture models along with some added complications necessary to convert abundance to density. Despite these difficulties, the "**`Density with Telemetry`**" data type in **MARK** provides a way forward and does so using auxiliary information that actually measures animal movement on and off of the study site, which is the source of the density estimation problem. There are other contemporary means of computing density as well, and we encourage readers to explore Ivan *et al*. (2013) for a discussion of advantages and disadvantages of this approach compared to others.

21.6. References

Anderson, D. R., Burnham, K. P., White, G. C., and Otis, D. L. (1983) Density estimation of small mammal populations using a trapping web design and distance sampling methods. *Ecology*, **64**, 674-680.

Borchers, D. L., and Efford, M. G. (2008) Spatially explicit maximum likelihood methods for capture-recapture studies. *Biometrics*, **64**, 377-385.

Converse, S. J., White, G. C., and Block, W. M. (2006) Small mammal responses to thinning and wildfire in ponderosa pine-dominated forests of the southwestern United States. *Journal of Wildlife Management*, **70**, 1711-1722.

Dice, L. R. (1938) Some census methods for mammals. *Journal of Wildlife Management*, **2**, 119-130.

Efford, M. G. (2004) Density estimation in live-trapping studies. *Oikos*, **106**, 598-610.

Efford, M. G., Borchers, D. L., and Byrom, A. E. (2009) Density estimation by spatially explicit capture-recapture: likelihood based methods. Pages 255-269 in D. L. Thomson, E. G. Cooch, and M. J. Conroy, editors. *Modeling Demographic Processes in Marked Populations*. Springer Science, Boston, Massachusetts, USA.

Huggins, R. M. (1989) On the statistical analysis of capture-recapture experiments. *Biometrika*, **76**, 133-140.

Huggins, R. M. (1991) Some practical aspects of a conditional likelihood approach to capture experiments. *Biometrics*, **47**, 725-732.

Ivan, J. S., White, G. C., and Shenk, T. M. (2013) Using simulation to compare methods for estimating density from capture-recapture data. *Ecology*, **94**, 817-826.

Kendall, W. L., Nichols, J. D., and Hines, J. E. (1997) Estimating temporary emigration using capture-recapture data with Pollock's robust design. *Ecology*, **78**, 563-578.

Link, W. A., and Barker, R. J. (1994) Density estimation using the trapping web design: a geometric analysis. *Biometrics*, **50**, 733-745.

Otis, D. L., Burnham, K. P., White, G. C., and Anderson D. R. (1978) Statistical inference from capture data on closed animal populations. *Wildlife Monographs*, 7, 135.

Parmenter, R. R., Yates, T. L., Anderson, D. R., Burnham, K. P., Dunnum, J. L., Franklin, A B., Friggens, M. T., Lubow, B. C., Miller, M., Olson, G. S., Parmenter, C. A., Pollard, J., Rexstad, E., Shenk, T., Stanley, T. R., and White, G. C. (2003) Small-mammal density estimation: a field comparison of grid-based vs. web-based density estimators. *Ecological Monographs*, **73**, 1-26.

Royle, J. A., and Dorazio, R. M. (2008) Conceptual and philosophical considerations in ecology and statistics. Pages 444 in J. A. Royle, and R. M. Dorazio, editors. *Hierarchical Modeling and Inference*

in Ecology. Academic Press, Boston, Massachusetts, USA.

Royle, J. A., Nichols, J. D., Karanth, K. U., and Gopalaswamy, A. M. (2009) A hierarchical model for estimating density in camera-trap studies. *Journal of Applied Ecology*, **46**, 118-127.

Schwarz, C. J., and Arnason, A. N. (1996) A general methodology for the analysis of capture-recapture experiments in open populations. *Biometrics*, **52**, 860-873.

White, G. C., and Shenk, T. M. (2001) Population estimation with radio-marked animals. Pages 329-350 in J. J. Millspaugh, and J. M. Marzluff, editors. *Radio tracking and animal populations.* Academic Press, San Diego, California, USA.

Williams, B. K., Nichols, J. D., and Conroy, M. J. (2002) *Analysis and Management of Animal Populations*. Academic Press, New York, New York.

Wilson, K. R., and Anderson, D. R. (1985) Evaluation of two density estimators of small mammal population size. *Journal of Mammalogy*, **66**, 13-21.

Zahratka, J. L., and Shenk, T. M. (2008) Population estimates of snowshoe hares in the southern Rocky Mountains. *Journal of Wildlife Management*, **72**, 906-912.