

Chapter 1

Spatial Mark-Resight Models

Throughout most of this book we have dealt with the situation where all individuals are identifiable upon encounter because they carry some form of individual mark. In Chapt. ?? we introduced and developed an SCR model for non-identifiable populations, a spatial *non-capture-recapture* model, if you will. These two extremes are common in the study of animal populations with non-invasive sampling methods. However, there is also an intermediate situation where part of the population is tagged or otherwise marked and can thus be identified upon recapture, while the unmarked portion remains unidentifiable. In this situation so-called mark-resight models (Bartmann et al., 1987; Arnsen et al., 1991; Neal et al., 1993) can be used to estimate population size and density by combining data from both the marked and unmarked individuals.

Traditionally, capture-recapture studies involved physical capture and marking of individuals throughout the study. This methodology is still widely applied in the study of species that are relatively easy to capture, such as small mammals, but can be very costly, logistically challenging and risky when dealing with larger species. In contrast, in mark-resight studies a sample of individuals is captured and tagged (or otherwise marked) during a single marking event. Marking is followed by resighting surveys, upon which both the detection of marked and unmarked animals is recorded. Resighting surveys are usually non-invasive (hence the name ‘resighting’), so that they don’t involve handling of animals. As such, mark-resight models have a major advantage over traditional capture-recapture models in that they only require individuals to be captured and handled once, during the initial marking. This reduces field costs and risks for the animals (and potentially the researchers).

Mark-resight models have a set of underlying assumptions, most of which are identical to those of capture-recapture models; e.g., demographic popula-

tion closure (violation of geographic population closure can be accommodated by some models) and no loss or misidentification of marks (see also Chapt. ??). Just like standard capture-recapture models, there are means to incorporate heterogeneity in capture probability. An essential assumption of mark-resight models is that the marked individuals are a representative sample of the study population, so that inference about detection can be made for the whole population from the marked sample. While this is also an implicit assumption of capture-recapture models, in mark-resight models this means that the process of marking individuals requires careful consideration in order to produce a random sample. This assumption is usually addressed by employing a different method for marking than for resighting.

Owing to the advantages of mark-resight over capture-recapture, especially when dealing with hard-to-trap species, mark-resight is a popular tool in wildlife population studies. The method has been applied for decades to a suite of species and survey techniques, ranging from banding and resighting Canada geese (Hestbeck and Malecki, 1989) to ear-tagging and camera-trapping grizzly bears (Mace et al., 1994) to paintball marking and areal resightings of large ungulates (Skalski et al., 2005).

In this chapter we consider mark-resight within a spatial context and develop a spatial mark-resight (SMR) model. To motivate this model development, imagine you conduct a live-trapping study during which you capture and mark a number of animals with individually recognizable tags. Subsequently, you go back out to the field and conduct resighting surveys on an array of locations, and during these resighting surveys you see some of your marked individuals, as well as new, unmarked ones. Then, for the marked animals you obtain the same type of spatially explicit individual encounter histories as you would in a standard SCR study. In addition, you obtain site (and occasion) specific counts of individuals you did not mark. SMR models make use of both the encounter history data from the marked individuals and the counts of unmarked individuals to estimate density and detection parameters.

In the following sections we first provide some background information on mark-resight and the types of data such surveys can provide. We will further explore the implications of the assumption of the marked individuals being a random subset of the population, which, in the context of SMR models refers to not only the *demographic composition*, but also to the *spatial distribution* of the marked individuals in the state-space \mathcal{S} . In many real life sampling situations, this assumption will not hold – animals will most often be marked in some region that does not represent the entire state-space. As a result, the distribution of marked individuals will generally *not* follow a homogeneous point process, but their activity centers will be concentrated in the vicinity of where marking took place. For the sake of model development, however, throughout the central part of this chapter we will make the assumption that marked animals are a random sample from the population in \mathcal{S} . We will show that SMR models are hybrids of standard SCR models and the models presented in Chapt. ?? for data where individuals cannot be uniquely identified. We explore models for both known and unknown numbers of marked individuals, and for imperfect individ-

ual identification of marks, and approaches to incorporate telemetry location data. In the spatial framework, most of the information on model parameters comes from the marked individuals. But in Sec. 1.5 we will see that, analogous to the models we developed previously in Chapt. ??, the spatial correlation in counts of unmarked individuals also contributes information about detection and movement. We conclude the chapter by presenting some general strategies for addressing a situation where marked individuals are not a random sample from \mathcal{S} .

1.1 Background

Before we start exploring spatial mark-resight approaches in more detail, we need to establish some terminology and gain a clear understanding of what types of mark-resight data we can have, in order to appreciate and understand the different flavors of mark-resight models.

1.1.1 Resighting techniques

As with capture-recapture surveys, there are numerous methods suitable for obtaining resightings. Common methods are visits to a set of points for resightings by an observer, or camera-trapping; but resightings need not be restricted to a particular set of locations. We can just as well envision a search-encounter kind of method, where a certain area is searched, systematically or opportunistically, for marked animals (see Chapt. ??). In this chapter we will only deal with fixed location resighting surveys, and we will refer to the set of resighting locations as the resighting array. In some instances we will also be concerned with where marked animals were captured, and we refer to these locations as the marking locations.

1.1.2 Types of mark-resighting data

In general, we have (at least) two sets of data: encounter histories for marked, and thus, identifiable individuals i at resighting location j and occasion k , y_{ijk} , and counts of unmarked records, n_{jk} , for each resighting location j and occasion k . Depending on the sampling technique, we can conceive of three slightly different types of partial ID data.

(1) Known number of marked individuals: If you implement a resighting survey shortly after the marking session, you may be confident that none of the marked individuals have died or lost their mark. Under these circumstances you know that the number of marked individuals available for resighting, m , is equal to the number of individuals you marked. Alternatively, the marking technique might involve radio-transmitters, allowing you to confirm the presence or absence of marked individuals in the resighting survey area using radio-telemetry (White and Shenk, 2001). In both cases, you know the number of marked individuals in the surveyed population. In this situation, even though you may fail

to resight some of the marked individuals, you know how many there are, and so you can simply assign all-zero encounter histories to the marked individuals not encountered – in other words, contrary to regular capture-recapture models, in mark-resight models with a known number of marked individuals, we can observe all-zero encounter histories. Under these circumstances, estimating N reduces to estimating the number of unmarked individuals, U .

(2) Unknown number of marked individuals: If m is not known, for example because we suspect that some of the marks may have been lost between tagging and conducting the resighting surveys, we obtain a slightly different type of mark-resight data. Here, we do not accurately know the number of marked individuals available for resighting. As a consequence, individuals have to be resighted at least once for us to know they are still marked and alive and thus available for resighting. So, contrary to the situation where we know m and analogous to regular capture-recapture models, we cannot observe all-zero encounter histories of marked individual. In this situation, estimating N involves estimating both m and U .

A special case of this kind of data can arise from camera trapping. Even when dealing with a species that has no spots or stripes, some individuals in the study population can have natural marks that make them identifiable on pictures, such as scars or a distinct coloration. In this scenario an individual has to be photographed at least once to be known. Here, the fact that both the “marking” method and the subsequent resighting method are the same (although marking in this case does not involve any actual physical marking) can be cause for concern: our sample of “marked” individuals may not be a random sample of the population but consist of individuals that for some reason are more likely to be photographed (e.g., individuals with activity center more interior to the trap array). In that case, a basic assumption of the mark-resight model is violated.

(3) Unknown marked status: Finally, consider a scat or hair snare survey, where only a part of the sample is analyzed genetically (or DNA can only be extracted from a subset of samples due to sample quality). In this scenario, n_{jk} can contain both completely unknown individuals that are not represented at all in the set of encounter histories of marked animals, Y , but it can also contain samples from individuals that we previously identified. The difference is that in the first two scenarios, part of the population of individuals is identifiable, while in the third scenario, part of the sample of individuals is identifiable. This type of data violates one of the basic assumptions of mark-resight models, namely, that marked individuals are always correctly identified as such.

To our knowledge there are currently no mark-resight models available that account for possible misidentification of the marking status of individuals (although some literature is available on misidentification of individuals in capture-recapture studies, e.g., Yoshizaki et al., 2009; Lukacs and Burnham, 2005; Link et al., 2010). In this chapter we will ignore this kind of data and focus instead on types (1) and (2).

For both types of data a slightly different situation arises when we can only tell that an individual is marked, but not who it is. You may be able to see that an individual is marked but the identifying feature of the tag (a number or coloration) may have become unreadable, or may be hidden from view. In this case, in addition to the observed y_{ijk} and n_{jk} , you also observe a number of sightings of marked but unidentified individuals, say r_{jk} .

1.1.3 A short history of mark-resight models

Initially, mark-resight methods focused on radio-tagged individuals to estimate population size (White and Shenk, 2001). Radio-collars provide a means of determining which of the animals are in the study area and available for sampling, thus determining the number of marked individuals in the population. Knowing this number was a prerequisite for most earlier mark-resight approaches (White, 1996). The oldest mark-resight model is the good old Lincoln-Petersen estimator, where individuals are marked and a single resight/recapture occasion is carried out (Krebs, 1999). We need not identify individuals, but only to tell apart marked from unmarked individuals. Let m be the number of marked individuals in the population, $m_{(R)}$ the number of marked individuals seen on the resighting occasion, and $n_{(R)}$ the total number of marked and unmarked individuals observed during resighting. Population size N is then estimated as

$$N = m \times n_{(R)} / m_{(R)}.$$

Mark-resight models using individual capture histories over several resighting occasions were developed in the 1980s and 90s and compiled into the program **NOREMARK** (White, 1996). Apart from the basic model with known number of marked individuals and no individual variation in resighting probabilities (joint hypergeometric maximum likelihood estimator) (Bartmann et al., 1987; White and Garrot, 1990; Neal, 1990; Neal et al., 1993), **NOREMARK** contains models that account for lack of geographic population closure (Neal et al., 1993), individual heterogeneity in resighting rates and sampling with replacement (i.e. individuals can be seen more than once on any occasion, (Minta and Mangel, 1989; Bowden, 1993)). A first mark-resight model allowing for an unknown number of marked individuals was developed by Arnason et al. (1991).

While many of these models perform well under certain situations, they are somewhat limited in that they do not allow for combining data across several surveys (McClintock et al., 2006) and not all of them are likelihood-based or allow for different parameterizations (e.g., including a time effect on detection), so that selection of the most appropriate model cannot be based on standard approaches such as AIC, but is largely left up to educated guesswork (McClintock et al., 2006). Recently, more flexible and generalized likelihood-based mark-resight models have been developed. These models can account for individual heterogeneity in detection, unknown number of marked individuals and lack of geographical closure, as well as a less than 100% individual identification rate of marked individuals; they can be applied to sampling with and without replacement and can combine data across several primary sampling occasions in

a **robust design type** of analysis (McClintock et al., 2009a,b). Since they are all likelihood-based, model selection among different parameterizations and model averaging based on AIC is an option. Most of these models have also been incorporated into the program **MARK** (McClintock and White, 2012).

For a detailed treatment of these different non-spatial mark-resight models, we refer you to the original papers cited in the preceding paragraph. In short, these models are based on the joint likelihood of two model components: one describing the resighting process of marked individuals and one describing the number of unmarked individuals observed. The resighting process of marked individuals can use either a Poisson or a Bernoulli observation model, depending on whether sampling is with or without replacement, and the resighting probabilities can have both fixed effects to model individual and environmental covariates, and a random-effect component to accommodate variation in detection due to individual heterogeneity. The process describing the number of unmarked individuals observed (or, under a Poisson observation model, the number of times unmarked individuals are observed), n_t (t here and in the following description denotes a primary sampling occasion, for example, a year or a season) is approximated as a normal distribution (McClintock et al., 2006), or a normal distribution left-truncated at 0 (McClintock et al., 2009a):

$$n_t \sim \text{Normal}(\mathbb{E}(n_t), \text{Var}(n_t)).$$

For a single-season study, the t subscript does not need to be included. Although this is a simplification of the actual sampling process, McClintock et al. (2006) found this normal distribution to be a satisfactory approximation, which allows N to enter the model likelihood via $\mathbb{E}(n_t)$ and $\text{Var}(n_t)$.

In the simplest model without any variation in detection, the expected number of resightings of unmarked individuals, $\mathbb{E}(n_t)$, can be written as the number of unmarked individuals times the expected number of detections of a single individual. This is the mean or expected value of the underlying observation model:

$$\mathbb{E}(n_t) = (N - m) * \theta \tag{1.1}$$

where $\theta = K \times p$ for a Binomial observation model with K replicates and individual detection probability p , or $\theta =$ expected individual encounter rate λ for a Poisson observation model. Similarly, $\text{Var}(n_t)$ depends on the underlying observation model and is based on the parameters that determine the individual detection probability/encounter rate. Combining these two components, N is directly incorporated into the joint likelihood of the model.

While these mark-resight models are very flexible, they share the shortcomings of traditional capture-recapture models when it comes to estimating population density (e.g., Chaps. ?? and ??). **As long as resightings are collected across a number of locations, however, they come with the same spatial information as (re)captures in a standard SCR study.** In the following sections we will consider mark-resight sampling in the framework of spatial capture-recapture.

1.1.4 The random sample assumption

In mark-resight studies it is a prerequisite that the marked portion of the studied population is a random sample of the population, so that detection probability for the population can be adequately estimated from the marked subset. If, for example, there is some latent group structure in the population where one group has a higher detection probability than the other, the marked portion of the population should have the same composition with regard to this group structure as the study population. Intuitively, people think of this as a demographic problem. But if you think back to Chapt. ?? and one of the motivations for the development of SCR models, this assumption also has spatial implications. In a non-spatial mark-resight study, if all the marked individuals live on the edge of the resighting array, their exposure to resighting will be lower compared to the exposure of unmarked individuals living in the center of the array, thus artificially deflating estimates of detection. So to obtain a truly random sample of the study population, the *locations of the home ranges* of the marked individuals also have to be a random sample of the home range locations of the entire population. In general, this will be difficult to assess or even to incorporate into study design or analysis, unless the spatial context of sampling is clearly defined. Thus, mark-resight models are, fundamentally, *spatial*.

In the SMR framework, this issue manifests itself more explicitly, for two reasons: (1) we define the spatial context of the population by setting a state-space; (2) we assume a certain distribution or point process for all individuals within that state-space, in most cases a uniform distribution or homogeneous point process (but see Chapt. ?? for models with inhomogeneous spatial point processes). For the marked individuals to follow a homogeneous point process (i.e., be a random spatial subset of the population) in \mathcal{S} , marking must be done uniformly throughout the state-space. When we study a species where some individuals can be identified based on natural marks, while others do not have unique marks (for example regular colored versus melanistic leopards), and we can assume that the distribution of these two groups of individuals across \mathcal{S} are identical, then we can frame the estimation problem in terms of estimating the density of two homogeneous point processes, one for the marked and one for the unmarked population. But what if we actively need to mark individuals in order to distinguish them? Then, we have two options: (1) if we want to meet the random sample assumption, then the definition of \mathcal{S} becomes part of our study design (contrary to SCR models, where \mathcal{S} is set after data collection for analysis purposes); (2) if we don't want to, or cannot, meet the random sample assumption, we have to specify an alternative model that adequately describes the distribution of marked and unmarked individuals in \mathcal{S} .

Here is another way to think about this: In SCR models, once the state-space is chosen large enough, estimates of density are no longer sensitive to the size of \mathcal{S} , because N scales with the area of \mathcal{S} . In spatial mark-resight, however, our population of individuals consists of two groups, marked and unmarked. Consider the case where we have a known number of marks. Because we fix the size of the marked part of our population, total population size N no longer

scales with the area of the state-space. While the number of unmarked individuals can go up as \mathcal{S} increases in size, m is fixed by design, and thus, as \mathcal{S} increases, overall density will decrease.

If we want to make sure by design that marked individuals are a random sample from \mathcal{S} , then, in practical terms, we need to define the state-space, which includes the resighting array plus a sufficient buffer to include all animals potentially exposed to this array, and uniformly mark individuals throughout \mathcal{S} . This does not mean that we necessarily have to achieve complete coverage of \mathcal{S} with our marking effort; alternatively, we could also randomly distribute traps in \mathcal{S} in order to randomly distribute marks throughout \mathcal{S} . We can see some sampling situations in which such a scenario might be reasonable, or at least reasonably approximated. For example, later on in this chapter we present a study where raccoons were caught and marked throughout an island, the boundaries of which are a natural limit for the state-space of this particular system.

For many studies, however, this might not be the case. Often, marking is the more difficult and logistically challenging part of a mark-resight study – think about the difficulty in physically capturing and tagging large carnivores. Especially for rare and cryptic species, areas over which resighting is conducted might have to be large to accumulate sufficient data, and marking over an even larger area – \mathcal{S} – would be logistically impossible. So what happens if we capture and mark individuals in a subset of the state-space? Then, whereas we may well have an overall constant density across \mathcal{S} , we will have a higher density of marked individuals in the vicinity of the marking locations – live traps, mist nets, whatever is used to catch animals – and the density of marks will generally go down as we get further away from the marking locations. As with all methods discussed in this book, the marking process of mark-resight studies also has a spatial component and induces a certain spatial distribution of marked individuals in the study area. We have to account for that when developing an SMR model. Thus, if we want to relax the assumption that marked animals are a random sample from \mathcal{S} , we need to describe the distribution of marked individuals' activity centers using an adequate spatial point process model. Developing a suitable point process model is one of the primary challenges when fitting SMR models, and one that at this point in time still requires substantial model development efforts. We provide some ideas on how to approach this problem in the last section of this chapter.

Although it might not be a reasonable assumption for many real life survey situations, for now, we will continue by showing the development of SMR models assuming that the marked animals are, indeed, a random sample of N , following a homogeneous point process in \mathcal{S} . This simplifies the modeling problem substantially, thereby allowing us to focus on the underlying principles and possible useful extensions of SMR models.

1.2 Known number of marked individuals

We begin the model development with the simplest situation. Here, a known number of individuals constituting a random sample from the population within \mathcal{S} are marked and a series of resight samples are conducted following marking. No marks (or marked animals) are lost between marking and resighting, all individuals are correctly identified as marked or unmarked, and marked individuals are 100% correctly identified to individual level.

Recall from Chapt. ?? that without any individual identity, the observed counts at resighting location j and occasion k , n_{jk} , represent the sum of all latent individual detections at j and k , $\sum_{i=1}^N y_{ijk}$, where y_{ijk} are the latent individual encounter histories. We can model these counts as

$$n_{jk} \sim \text{Poisson}(\Lambda_j)$$

where

$$\Lambda_j = \sum_{i=1}^M (\lambda_{ij}).$$

Under this formulation, in order to carry out MCMC, we do not need to update the individual y_{ijk} in our model, which is more efficient in terms of computing. However, we can also formulate the model as conditional on the latent y_{ijk} . This is useful because if we have m marked animals in our study population, then y_{ijk} for those m individuals are no longer latent, but fully observed and can easily be included in the analysis to provide information on detection parameters.

The formulation conditional on y_{ijk} basically brings us back to the original SCR model, where individual site and occasion specific counts, y_{ijk} , are modeled as

$$y_{ijk} \sim \text{Poisson}(\lambda_{ij})$$

and

$$\lambda_{ij} = \lambda_0 \exp(-d_{ij}^2 / (2\sigma^2)).$$

in the case of a Poisson encounter model. Unobserved y_{ijk} are treated as missing data and have to be updated as part of the MCMC procedure. We can do that by using their full conditional distribution, which is multinomial with sample size n_{jk} . Let \mathbf{u} be an index vector of the $M - m$ hypothetical unmarked individuals, $u = m + 1, m + 2, \dots, M$, and let \mathbf{y}_{ujk} be the vector of observations of all individuals in \mathbf{u} at j and k . Then

$$\mathbf{y}_{ujk} \sim \text{Multinomial}(n_{jk}, \boldsymbol{\lambda}_{uj})$$

Whereas in the non-spatial mark-resight analysis, known individuals provide information about individual detection probability (or rate), in the spatial setting they also inform σ , as described in Chapt. ?. Including known individuals into the analysis helps estimate model parameters more accurately and precisely. We will address the relationship between the number of marked individuals and accuracy of the estimated parameters in Sec. 1.5.

1.2.1 Implementing spatial mark-resight models

Implementing a spatial mark-resight model in **JAGS** is not straightforward, since the program does not accept partially observed multivariate nodes (in this case the partially observed individual encounter histories which we model as coming from a multinomial distribution). We can, however, work around that by separating the marked from the unmarked data. The **JAGS** code for the model with a known number of marked individuals is shown in Panel 1.1. You see that data augmentation is only applied to the unmarked part of the population, and N is the sum of the estimated number of unmarked individuals (`sum(z[])`) and the number of marked individuals, which is known. Also, to reduce run time, we summed observations of marked individuals across occasions and account for that by multiplying λ_{ij} with K . Although the two data sets are separated, both parts of the population, marked and unmarked, have the same prior uniform distribution of activity centers. A last noteworthy detail in this code is the `dsum()` distribution. This distribution is specific to **JAGS** (i.e., you cannot run this model in **BUGS**), and allows you to impose a sum constraint on observations. In other words, it allows to model data – here the counts of unmarked individuals, n_{jk} – as the sum of a number of latent variables, which in this case are the latent encounter histories of unmarked individuals. While it can be a pain writing out all the arguments of `dsum()`, it is this function that allows us to implement SMR models in **JAGS**.

Alternatively, we can use the technical concepts presented in Chapt. ?? and derive our own MCMC algorithm. To do so, we only have to make relatively simple modifications to the MCMC code developed for regular SCR models in Chapt. ?. Essentially, since we observe individual detections for the marked part of the population, we have to update only the unobserved part of the full – augmented – set of encounter histories, \mathbf{Y} , and modify the updating steps for z_i and ψ , the parameters introduced by data augmentation, to reflect that these only apply to the unmarked part of the population, in other words, to the $M - m$ individuals in our data. You can find the full MCMC code in the accompanying **R** package `scrbook` by invoking `scrPID`. The **R** code below shows how to simulate SMR data using the `scrbook` function `sim.pID.data`, and running an SMR model on the data, both in **JAGS** and using `scrPID`. The model file `mknown.jag` in the `jags.model` call should contain the code from Panel 1.1.

```

397 > set.seed(2013)
398 > N = 80 # pop. size
399 > m <- 45 # no. marked
400 > sigma = 0.5
401 > lam0 = 0.5
402 > K = 5
403 # Make resighting array
404 > gx <- gy <- seq(0,6,1)
405 > X <- as.matrix(expand.grid(gx, gy))
406 > J = dim(X)[1]

```

```

model{

#priors
psi ~ dbeta(1,1)
lam0 ~ dunif(0, 5)
sigma ~ dunif(0, 5)

#marked part
for(i in 1:m) {
  sm[i,1] ~ dunif(xlim[1], xlim[2])
  sm[i,2] ~ dunif(ylim[1], ylim[2])
  for(j in 1:J) {
    distm[i,j] <- sqrt((sm[i,1]-X[j,1])^2 + (sm[i,2]-X[j,2])^2)
    lambdam[i,j] <- lam0*exp(-distm[i,j]^2/(2*sigma^2))
    y[i,j]~dpois(lambdam[i,j]*K)
  }
}

##unmarked part
for(i in 1:M) {
  z[i] ~ dbern(psi)
  s[i,1] ~ dunif(xlim[1], xlim[2])
  s[i,2] ~ dunif(ylim[1], ylim[2])
  for(j in 1:J) {
    dist[i,j] <- sqrt((s[i,1]-X[j,1])^2 + (s[i,2]-X[j,2])^2)
    lambda[i,j] <- lam0*exp(-dist[i,j]^2/(2*sigma^2))
    for(k in 1:K) {
      yu[i,j,k] ~ dpois(lambda[i,j]*z[i])
    }
  }
}

for(j in 1:J) {
  for(k in 1:K) {nU[j,k] ~ dsum(yu[1,j,k],yu[2,j,k],yu[3,j,k],
[...code shortened...],
yu[79,j,k],yu[80,j,k])
}
}

N <- sum(z[])+m

}

```

Panel 1.1: **JAGS** model specification for SMR model with known number of marked individuals. In this example, M , the size of the augmented unmarked data set, is 80. Note that the arguments $yu[4,j,k]$ to $yu[78,j,k]$ of the `dsum()` function are omitted from the code.

```

407 > # Limits of S
408 > xlims <- ylims<-c(-1.5, 7.5)
409 # Simulate data
410 > dat = sim.pID.data(N=N, K=K, sigma=sigma, lam0=lam0, knownID=m, X=X, xlims=xlims,
411 ylims=ylims, obsmod='pois',nmarked='known')
412
413 ### Prep data for analysis in JAGS
414 > n <- dat$n - apply(dat$Yknown,2:3,sum)
415 > y <- apply(dat$Yknown,1:2,sum)
416
417 > M <- 80 # Augmentation only for unmarked
418
419 # Initial values for latent y
420 > yin <- array(0, c(M,J,K))
421 > for(j in 1:J){
422 +   for(k in 1:K){
423 +     yin[1:M,j,k] <- rmultinom(1, n[j,k], rep(1/M, M))
424 +   }}
425
426 > data <- list(y=y, nU=n, m=m, M=M, J=J, X=X, xlim=xlims, ylim=ylims, K=K)
427 > inits <- function(){list(sigma=runif(1), lam0=runif(1),
428 sm=cbind(runif(m, xlims[1], xlims[2]), runif(m, ylims[1], ylims[2])),
429 s=cbind(runif(M, xlims[1], xlims[2]), runif(M, ylims[1], ylims[2])),
430 z=rep(1, M),yu=yin)}
431 > params <- c('lam0', 'sigma', 'N', 'psi')
432
433 # Analysis in JAGS
434 > library(rjags)
435 > mod <- jags.model('mknown.jag',data, inits, n.chains=1, n.adapt=800)
436 > out <- coda.samples(mod,params, n.iter=5000)
437
438 > # Analysis with scrbook MCMC code
439 > library(scrbook)
440 > library(coda)
441 > inits2 <- function(){list(psi=runif(1), sigma=0.5, lam0=0.5,
442 S=cbind(runif(M+m, xlims[1],xlims[2] ), runif(M+m, ylims[1],ylims[2])))}
443 > out2 <- scrPID(n=n, X=X, y=dat$Yknown, M=M+m, obsmod = "pois", niters=5800,
444 xlims=xlims, ylims=ylims,inits=inits2(),delta=c(0.1,0.1,0.5))

```

You can look at the two sets of output by invoking `summary(out)` for the **JAGS** analysis and `summary(window(mcmc(out2),start=801))` for the custom MCMC algorithm, excluding the first 800 iterations as burn-in. We summarized the results in Table 1.2.1. The posterior mean of N is slightly higher than the data-generating value of $N = 80$, but it falls comfortably within the credible intervals. As expected, estimates from both implementations are very similar; slight differences are probably the result of Monte Carlo error due to the relatively low number of iterations. You will find that sometimes, **JAGS** produces an error message upon trying to compile the model, saying that some of the observed y are inconsistent with parent nodes at initialization. We have

Table 1.1: Posterior summaries of the parameters of a spatial mark-resight model with known number of marks, analyzed in JAGS and using `scrPID`.

Implementation	Parameter	Mean	SD	2.5%	50%	97.5%
JAGS	N	88.72	6.75	77	88	103
	λ_0	0.53	0.08	0.39	0.53	70
	σ	1.29	0.02	1.26	1.30	1.32
	ψ	0.47	0.03	0.45	0.47	0.53
scrPID	N	86.01	7.58	73	85	102
	λ_0	0.54	0.08	0.39	0.53	0.72
	σ	0.48	0.03	0.42	0.48	0.53
	ψ	0.51	0.11	0.32	0.51	0.73

mentioned before that **JAGS** cannot always auto-generate acceptable initial values, and we believe this is what is happening here. If this error occurs, just repeat the `jags.model` command, usually, model compilation is successful on a second attempt (assuming, of course, that you followed the code above correctly). We further find that the custom MCMC algorithm tends to be faster than **JAGS**, which is why the examples and simulation studies shown in the following sections were run solely in **R**.

1.3 Unknown number of marked individuals

Now let us consider the case where we do *not* know the exact number of marked individuals available for resighting so that we have to capture an individual at least once to be sure that it is available. Unless we have a direct means of confirming the number of marked animals available for resighting, treating this number as unknown is probably more realistic in most circumstances. As a consequence of not knowing the exact number of marked individuals, we cannot observe all-zero encounter histories. When using maximum likelihood inference, this situation requires a model where detection rates of known individuals are modeled using a zero-truncated distribution (McClintock et al., 2009a). If we did not account for the fact that zeros are unobservable, estimates of detection rates would be artificially inflated and estimates of population size would be negatively biased.

Working with zero-truncated distributions in a spatial mark-resight setting is less straight-forward than for non-spatial mark-resight. A marked individual only has to show up once, anywhere on the resighting array, for us to know that it is there. When resightings are pooled across the entire sampling grid, then the total individual counts $\sum_j y_{ijk}$ have to be > 0 for all resighted individuals and a zero-truncated distribution can be used to model these counts. However, we are concerned with trap-specific encounters, y_{ijk} , which can easily be 0 for a resighted individual, as long as a single y_{ij} is > 0 . Thus, the zero-truncation does not apply to the individual and trap specific counts we observe, but only to the sum of these counts over all traps.

As an alternative to a zero-truncated distribution, in a Bayesian framework, we can make use of data augmentation to estimate the number of marked individuals (McClintock and Hoeting, 2010). In the SMR framework that means that we create two augmented data sets, one for the marked individuals and one for the unmarked, and estimate their number separately, having them share the parameters of the detection model. Sometimes we may know the maximum number that were ever marked before a resighting survey, in which case we can use that number as the data augmentation limit for the marked data set. Panel 1.2 shows the **JAGS** code for the SMR model with unknown number of marks, which is identical to the one in Panel 1.1, but for the augmentation of the marked data set. This introduces both a data augmentation parameter, `psim`, and an auxiliary “alive state” variable, `zm[i]`, into the description of the marked data model. Again, we provide an alternative, **R**-only MCMC algorithm within `scrbook` – `scrPID.um`.

Note that we could look at the problem of not knowing the number of marked individuals in the study population as a manifestation of a lack of population closure. In other words, marked individuals may have emigrated, died or lost their marks in the time between marking and resighting. If we have information on the rates of these events, or a series of resighting surveys, we could develop an open population model for the marks in our population and estimate their number at a given resighting survey in this fashion. This kind of SMR model remains to be explored.

1.3.1 Canada geese in North Carolina

We applied the spatial mark-resight model with an unknown number of marks and a binomial encounter process to a dataset of Canada goose resightings (Rutledge, 2013). During the molt of 2008, 751 individual geese were captured and marked with neck and leg bands in Greensboro, North Carolina (Fig. 1.1). Geese were resighted at 87 locations on 81 resighting events over a period of 18 months. In addition to the banded geese, the number of unmarked geese was recorded during each resighting event. Here, we only looked at a subset of the data, from mid July to the end of October 2008, which corresponds to the first part of the post-molt season, before migratory Canada geese arrive in North Carolina. We treated this population as closed over this period. During this part of the study, 57 of the resighting sites were visited and $n = 654$ marked geese were resighted 3994 times at 40 different sites. In addition, 7944 sightings of unmarked geese were recorded at 48 sites.

In the model, we allowed σ to vary between males and females. We set the size of the augmented unmarked data set to 7000. We used the total number of marked geese (751) as the upper limit for the augmented marked data set. We ran 50000 MCMC iterations and removed a burn-in of 5000 iterations. To describe the state-space, we buffered the resighting locations by 4.5 km. We assumed that marked geese were a random sample from the state-space, which seems reasonable because (a) marking took place across most of the extent of the resighting array; and (b) marking was done during the molting period, when

```

model{

# Prior distributions
psim ~ dbeta(1,1)
psi ~ dbeta(1,1)
lam0 ~ dunif(0, 5)
sigma ~ dunif(0, 5)

# Marked part of the model
for(i in 1:max) {
  zm[i]~dbern(psim)
  sm[i,1] ~ dunif(xlim[1], xlim[2])
  sm[i,2] ~ dunif(ylim[1], ylim[2])
  for(j in 1:J) {
    distm[i,j] <- sqrt((sm[i,1]-X[j,1])^2 + (sm[i,2]-X[j,2])^2)
    lambdam[i,j] <- lam0*exp(-distm[i,j]^2/(2*sigma^2))*zm[i]
    y[i,j]~dpois(lambdam[i,j]*K*z[i])
  }
}

# Unmarked part of the model
for(i in 1:M) {
  z[i] ~ dbern(psi)
  s[i,1] ~ dunif(xlim[1], xlim[2])
  s[i,2] ~ dunif(ylim[1], ylim[2])
  for(j in 1:J) {
    dist[i,j] <- sqrt((s[i,1]-X[j,1])^2 + (s[i,2]-X[j,2])^2)
    lambda[i,j] <- lam0*exp(-dist[i,j]^2/(2*sigma^2))
    for(k in 1:K) {
      yu[i,j,k] ~ dpois(lambda[i,j]*z[i])
    }
  }
}

for(j in 1:J) {
  for(k in 1:K) {nU[j,k] ~ dsum(yu[1,j,k],yu[2,j,k],yu[3,j,k],
[...code shortened...],
yu[79,j,k],yu[80,j,k])
}
}

Nu <- sum(z[])
Nm<-sum(zm[])
N<-Nu+Nm

}

```

Panel 1.2: JAGS model specification for SMR model with unknown number of marked individuals. In this example, M , the size of the augmented unmarked data set, is 80. Note that the arguments $yu[4,j,k]$ to $yu[78,j,k]$ of the `dsum()` function are omitted from the code for space reasons.

Table 1.2: Posterior summaries of parameters of the spatial mark-resight model for Canada geese in North Carolina. N is the total population size of marked and unmarked individuals; m is the number of marked individuals.

	Mean	SD	2.5%	50%	97.5%
m	739.77	3.24	733	740	746
N	5756.10	90.68	5577	5757	5932
D	13.76	0.19	13.38	13.76	14.14
λ_0	0.19	<0.01	0.18	0.19	0.19
σ , females	1.29	0.02	1.26	1.30	1.32
σ , males	1.06	0.02	1.02	1.06	1.11
ψ , marked	0.99	<0.01	0.98	0.99	0.99
ψ , unmarked	0.72	0.01	0.69	0.72	0.74
ϕ	0.36	0.02	0.32	0.36	0.39

geese are fairly immobile, and it seems reasonable to assume that, once the molt is complete, the marked geese redistributed themselves. Under this model formulation, estimates of density will be sensitive to the choice of the state-space (1.1.4), but the particular state-space seems like a reasonable choice for this problem. We provide all the data (`data('geesedata')`) and functions (`geeseSMR`) for you to repeat this analysis but be aware that given the large data set it will take days to do so. The **R** code to set up the data and run 5000 iterations of the model for the geese data is given as an example on the help page for `geeseSMR`. The model results, including the derived parameter density (D) in individuals per km^2 are shown in Table 1.3.1.

We see that credible intervals of estimates are pretty narrow, surely an effect of the large data set. Estimates of m indicate that most of the 751 geese originally banded are still alive and marked, which is not surprising, given that not much time passed between marking and this first resighting session. The parameter ϕ in this model is the probability of being a male, a measure of the sex ratio of the population, which is slightly biased in favor of females.

1.4 Imperfect identification of marked individuals

Often during resighting, it may be possible to see that an individual is marked but impossible to determine its individual identity. In this situation, in addition to y_{ijk} and n_{jk} , we also have site and occasion specific counts of marked but unidentified individuals, r_{jk} . Here, the individual encounter histories of marked animals are incomplete, and if we used these incomplete data to inform the detection parameter of the model, we would run the risk of underestimating encounter rate and overestimating abundance. Some non-spatial mark-resight models do not require that marked animals be identified individually, as long as the marking status can be observed unambiguously, but ignoring individual



Figure 1.1: Banded and unbanded Canada geese in a parking lot in Greensboro, North Carolina. (Photo credit: M.E. Rutledge, NCSU Canada Goose Project)

level information means that we cannot accommodate heterogeneity in detection (McClintock and White, 2012). In a spatial framework we could ignore marked and unmarked status completely and apply the model by Chandler and Royle (2013) discussed in Chapt. ???. But, that would mean losing important information on individual detection and movement. Therefore, being able to retain the individual identity of records that can be identified while at the same time accounting for imperfect identification of marked individuals is extremely useful.

McClintock et al. (2009a,b) suggest an intuitive means of correcting for this bias in a non-spatial model framework when dealing with a Poisson encounter model (a plausible model when sampling with replacement). When marked but unknown resightings are part of the data, the expected number of unmarked records, n , changes from Eq. 1.1.3 to:

$$\mathbb{E}(n) = (N - m)\lambda + \eta/m$$

where λ is the individual encounter rate estimated from the known resighted individuals and η is the number of records of marked but unidentified individu-

als. So, because the observed λ is known to be too low, the average number of unidentified pictures per known individual is added as a correction factor. This procedure assumes that the inability to identify a marked individual occurs at random throughout the population, which seems to be a reasonable assumption under most circumstances.

We can translate this same concept to the spatial mark-resight models. In the spatial framework we are interested in the individual and trap specific encounter rate, λ_{ij} . Further, we do not look at the sum of all records of unmarked individuals, but formulate the model conditional on the latent individual encounter histories. Thus, instead of using η/m as a correction factor, we need something that applies at the individual and trap level. If we take the sum of all correctly identified records of marked individuals, $\sum y_c$ and divide it by the total number of records of marked individuals, $\sum y_m$, we get the average rate of correct individual identification for marked individuals, say, c :

$$c = \sum y_c / \sum y_m.$$

We can then apply c as a correction factor for λ_0 for the marked individuals.

A more formal, model-based way to specify c is by assuming that

$$\sum y_c \sim \text{Binomial}(\sum y_m, c)$$

and estimating c as another model parameter, so that we account for the uncertainty about it. For the marked individuals we can then multiply λ_0 by c to account for the fact that we observe incomplete individual encounter histories. Since we don't have this identification issue for unmarked individuals, their baseline trap encounter rate remains as before simply λ_0 (or in other words, c for unmarked individuals equals 1).

Incomplete individual identification of marked individuals is easily incorporated into our **JAGS** model, no matter whether m known or unknown, by adding the following two lines of code:

```
c ~ dbeta(1,1) #prior for c
npics[1] ~ dbin(c, npics[2]) #model for c
```

and modifying the marked observation model description to

```
y[i,j] ~ dpois(lambdam[i,j]*c*K)
```

Here, the data object `npics` is a vector with the number of correctly identified records of marked individuals and the total number of marked records. Accounting for imperfect identification of marks is also included as an option in the `scrPID` and `scrPID.um` functions. Choosing an uninformative (and conjugate) $\text{beta}(1, 1)$ prior for c , within the `scrPID` algorithm we can update c directly from its full conditional distribution, which is $\text{beta}(1 + \sum y_c, 1 + (\sum y_m - \sum y_c))$. We show an example of using c in an analysis in Sec. 1.6.

Observe that now, in addition to assuming that failure to identify marked individuals occurs at random throughout the population, we also assume that

it occurs at random throughout space, i.e. our success of identifying a marked individual does not depend on the trap we encounter it in. As long as individuals are identified based on the same type of tags the assumption that failure to identify marked individuals occurs at random throughout the population should be valid. The assumption that failure to identify marked individuals occurs at random in space could be violated, for example when spatially varying habitat conditions influence the ability to recognize individual tags, or when an observer effect influences individual identification rates. While we haven't experimented with it, we believe that the approach described above could readily be extended to account for these differences. For example, identification rates could be calculated separately for different observers, or be modeled as functions of habitat covariates. As an alternative to the approach we present here, model development could explore assigning records of marked but unidentified individuals to marked individuals in a fashion similar to how unmarked records are assigned to hypothetical individuals in this model, namely, based on the location of the record and the estimates of home range centers of marked individuals. While this is computationally more advanced it would make full use of the spatial information of the unmarked records.

1.5 How Much Information Do Marked and Unmarked Individuals Contribute?

It is intuitive that having marked individuals in the study population should lead to more accurate and precise parameter estimates than when no individuals are identifiable. To evaluate how strongly adding marked individuals to a population improves parameter estimates, Chandler and Royle (2013) performed a simulation study. They used a 15×15 resighting grid and simulated detection data of $N = 75$ individuals in a 20×20 units state-space over $K = 5$ occasions with $\sigma = 0.5$ and $\lambda_0 = 0.5$. They generated 100 datasets each for $m = (0, 5, 15, 25, 35)$ where m is the known number of marked individuals randomly sampled from the population.

Without any marked individuals in the population, the posterior distribution of N turned out to be highly skewed, but the mode was still an approximately (frequentist) unbiased point estimator of N . As anticipated, posterior precision increased substantially with the proportion of marked individuals (Table 1.3 and Fig. 1.2). The relative root-mean squared error decreased from 0.246 when no individuals were marked to 0.085 when 35 individuals were marked (Table 1.3). Coverage was nominal for all values of m and posterior skew greatly diminished with increasing m (Table 1.3).

As we saw in the previous chapter, the spatial correlation in unmarked counts can be sufficient to obtain estimates of movement and detection parameters. However, only marked and thus identifiable individuals provide us with direct information about these parameters and may well dominate estimates. To single out the contribution of marked and unmarked individuals to parameter esti-

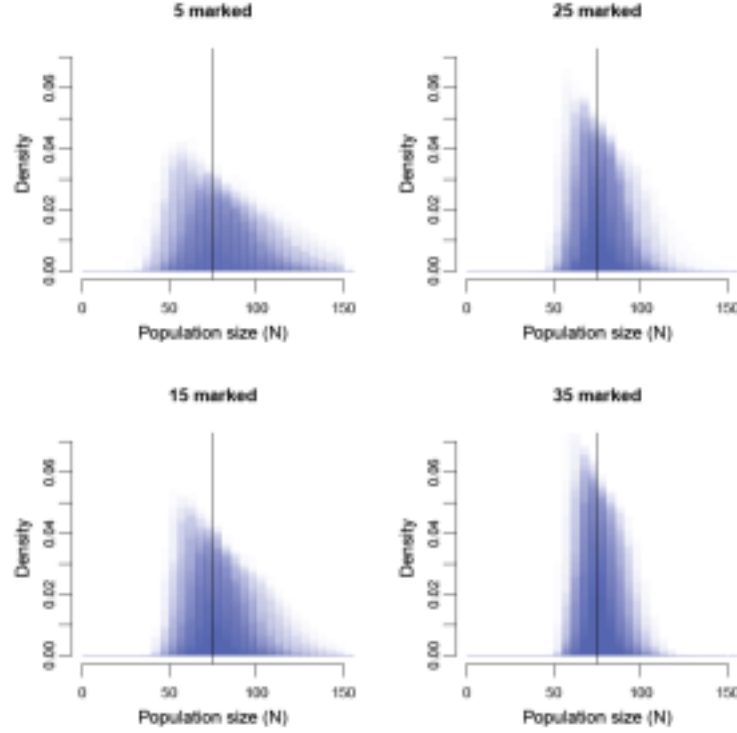


Figure 1.2: Overlaid posterior distributions of N from 100 simulations for four levels of marked individuals.

651 mates, we re-ran the same simulations but let σ and λ_0 be updated based solely
 652 on the data of marked individuals. Results are summarized in Table 1.4. We see
 653 that if we update λ_0 and σ based on marked individuals only, estimates of these
 654 parameters are more biased and less precise. For estimates of N , especially for
 655 $m=5$ and $m=15$, we observe a stronger positive bias, lower accuracy and con-
 656 siderably lower BCI coverage as compared to when both marked and unmarked
 657 individuals contribute to parameter estimates (Table 1.4). Thus, unmarked
 658 individuals do actually contribute noticeably to estimating model parameters.
 659 This stands in contrast to non-spatial mark-resight models, where information
 660 about the detection parameter comes solely from the encounter histories of the
 661 marked individuals.

662 1.6 Incorporating telemetry data

663 As we expected, parameter estimates of spatial mark-resight models get better
 664 the more marked individuals we have in our study population. While this is

Table 1.3: Posterior mean, mode, and associated relative RMSE for simulations in which m of $N=75$ individuals were marked. One hundred simulations of each case were conducted. Table taken from Chandler and Royle (2013)

	Parameter	Mean	rRMSE	Mode	rRMSE	BCI
m=0	N	85.866	0.259	77.720	0.242	0.950
	λ_0	0.506	0.180	0.488	0.182	0.960
	σ	0.495	0.115	0.486	0.113	0.960
m=5	N	80.898	0.184	76.360	0.182	0.970
	λ_0	0.510	0.178	0.494	0.180	0.950
	σ	0.496	0.089	0.488	0.086	0.970
m=15	N	79.028	0.148	76.250	0.147	0.950
	λ_0	0.508	0.163	0.494	0.164	0.950
	σ	0.496	0.073	0.492	0.071	0.970
m=25	N	77.765	0.114	75.810	0.113	0.950
	λ_0	0.511	0.153	0.498	0.157	0.950
	σ	0.496	0.067	0.493	0.065	0.940
m=35	N	76.446	0.085	74.900	0.085	1.000
	λ_0	0.513	0.142	0.501	0.144	0.950
	σ	0.497	0.056	0.493	0.057	0.940

great advice in theory, it may not be very helpful in practice, especially when dealing with animals that are hard or somewhat dangerous to capture, such as large carnivores. Oftentimes, studies involving the physical capture of such animals will employ telemetry tags in order to learn about the study species' spatial ecology and behavior. In the context of spatial mark-resight models, the actual locations collected by telemetry tags can provide detailed information on individual location and movement, and being able to incorporate this information directly into the SMR model should improve estimates of these parameters, especially when resighting information is sparse.

So how could we combine resighting data and telemetry data in a unified mark-resight model? Recall that the basic SCR model underlying all the SMR models we discuss here uses a Gaussian kernel to describe the trap encounter model. By using this function, we can relate the parameters σ and \mathbf{s}_i directly to those from a bivariate normal model of space usage, with mean = \mathbf{s}_i , and variance-covariance matrix Σ , where the variance in both dimensions is σ^2 and the covariance is 0. Ordinarily, these parameters are estimated directly from the spatial distribution of individual captures/resightings. Telemetry data, however, provide more detailed information on individual location and movement, since the resolution and extent of the data are not limited by the trapping grid and potentially more locations can be accumulated through telemetry than resighting (depending on the monitoring frequency and resighting rates of individuals).

By assuming that the R_i locations of individual i , \mathbf{I}_i (consisting of a pair of x

Table 1.4: Posterior mean, mode, and associated relative RMSE for simulations in which m of $N=75$ individuals were marked and unmarked individuals did not contribute to estimating λ_0 and σ . One hundred simulations of each case were conducted.

	Parameter	Mean	RMSE	Mode	RMSE	BCI
m=5	N	88.621	0.369	83.139	0.421	0.810
	λ_0	1.255	1.247	0.606	1.148	0.950
	σ	0.472	0.252	0.426	0.333	0.910
m=15	N	81.031	0.192	78.361	0.175	0.820
	λ_0	0.535	0.281	0.476	0.284	0.970
	σ	0.503	0.109	0.490	0.107	0.940
m=25	N	78.206	0.129	76.594	0.123	0.920
	λ_0	0.531	0.204	0.496	0.202	0.960
	σ	0.497	0.081	0.489	0.084	0.950
m=35	N	76.833	0.099	75.422	0.096	0.940
	λ_0	0.528	0.192	0.505	0.186	0.940
	σ	0.499	0.069	0.493	0.070	0.960

and y coordinates, l_{ix} and l_{iy}), are a bivariate normal (BVN) random variable:

$$\mathbf{l}_i \sim \text{BVN}(\mathbf{s}_i, \Sigma)$$

we can estimate σ as well as \mathbf{s}_i for the collared individuals directly from telemetry locations, using their full conditional distributions:

$$[\sigma | \mathbf{l}, \mathbf{s}] \propto \left\{ \prod_{i=1}^m \prod_{r=1}^{R_i} \frac{1}{2\pi\sigma^2} \exp \left(-1/2 \left[\frac{(l_{irx} - s_{ix})^2}{\sigma^2} + \frac{(l_{iry} - s_{iy})^2}{\sigma^2} \right] \right) \right\} * [\sigma]$$

and

$$[\mathbf{s}_i | \mathbf{l}, \sigma] \propto \left\{ \prod_{r=1}^{R_i} \frac{1}{2\pi\sigma^2} \exp \left(-1/2 \left[\frac{(l_{irx} - s_{ix})^2}{\sigma^2} + \frac{(l_{iry} - s_{iy})^2}{\sigma^2} \right] \right) \right\} * [\mathbf{s}_i]$$

For the unmarked individuals \mathbf{s}_i are estimated as described before conditional on their latent encounter histories. Note that the bivariate normal model assumes that locations are independent of each other. If you have frequent telemetry fixes, for example from GPS collars that report animal locations every few hours or more, this assumption seems unrealistic and it might be advisable to thin your telemetry data (maybe to daily fixes) in order to approximate independence. Alternatively, movement models could be used that acknowledge the temporal correlation in location data. We suggested some possible movement models in Chapt. ???. Not all marked individuals need to be telemetry-tagged, but telemetry data should correspond to the period over which resighting surveys were conducted (as we discussed in Chapt. ??, both the \mathbf{s}_i and σ should only be interpreted against the specific sampling period). Further, telemetry data need to be independent of the resighting data.

Again, implementation of this model extension is straight-forward, both in **JAGS** and **R**. Take the SMR model description for the case where m is known (Panel 1.1). Then, all we have to do is add a description of the bivariate normal model for the telemetry locations, here `locs`, into the loop over the m marked individuals:

```
[...parts of model code omitted...]

for(i in 1:m) {
  sm[i,1] ~ dunif(xlim[1], xlim[2])
  sm[i,2] ~ dunif(ylim[1], ylim[2])

  #telemetry model
  for (r in off1[i]:off2[i]){
    locs[r,1]~dnorm(sm[i,1], 1/(sigma^2))
    locs[r,2]~dnorm(sm[i,2], 1/(sigma^2))
  }

  for(j in 1:J) {
    distm[i,j] <- sqrt((sm[i,1]-X[j,1])^2 + (sm[i,2]-X[j,2])^2)
    lambdam[i,j] <- lam0*exp(-distm[i,j]^2/(2*sigma^2))
    y[i,j]~dpois(lambdam[i,j]*K)
  }
}

[...parts of model code omitted...]
```

The data object `locs` is a table with all $\sum_i^m R_i$ telemetry locations. The two vectors `off1` and `off2` describe which subset of this matrix belongs to individual i . So if, say, the locations for individual 1 are contained in the first 10 rows of `locs`, `off1` and `off2` would be 1 and 10 for $i = 1$; and if the locations of individual 2 are in the following 15 rows, `off1` and `off2` for $i = 2$ would be 11 and 25, and so on. For the implementation of this SMR model with telemetry data in **R**, see the `scrPID.tel` function in `scrbook`. In a nutshell, in the MCMC algorithm we replaced the Metropolis-Hastings updating steps for σ and activity centers of marked individuals, which were originally conditional on the resighting data, with updating steps conditional on the telemetry data. This is not quite what the above **JAGS** code does; rather **JAGS** will update these parameters conditional on both the telemetry *and* the resighting data. We could easily re-write `scrPID.tel` to do that, but believe that for most applications, the information on location and movement contained in the telemetry data will outweigh that in the resighting data, so that the resulting loss of information should be minimal.

1.6.1 Raccoons on the Outer Banks of North Carolina

Sollmann et al. (2013) applied a spatial mark-resight model with telemetry data to a camera-trap and radio-telemetry data set from the raccoon population on

749 South Core Banks, a barrier island within Cape Lookout National Seashore,
 750 North Carolina. Between May and September 2007, 131 raccoons were marked
 751 with dog collars and large individually numbered cattle tags. Individuals were
 752 marked throughout the island, so that (a) we do not have to deal with sensitivity
 753 to choice of the state-space, because it is clearly defined by nature; and (b) it is
 754 reasonable to assume that marked raccoons are a random sample of individuals
 755 from this state-space. Of the 131 tagged individuals, 44 were also equipped
 756 with radio collars. Collared individuals were located using a VHF receiver and
 757 antenna, and their locations were estimated approximately weekly. Twenty
 758 camera traps were set up along the length of South Core Banks and camera
 759 trapping data collected between October 1 2007 to January 22 2008 constituted
 760 the resighting data in this analysis. During this period 104 marked individuals,
 761 38 radio-collared, were alive and available for resighting with camera traps.



Figure 1.3: Camera trap picture of a raccoon marked with a cattle tag that cannot be read to determine individual identity. Taken on South Core Banks, North Carolina. (*Photo credit: Arielle Parsons*)

762 The state-space \mathcal{S} was the entire area of South Core Banks island. A change
 763 in the number of photocaptures over the course of the study suggested a varia-
 764 tion of detection rate with time. Since date recording in cameras malfunctioned,
 765 photographic records could only be assigned to the time interval between subse-
 766 quent trap checks, and these intervals between checks are referred to as sampling
 767 occasions. These occasions ranged from 2 to 43 days; λ_0 was standardized to
 768 7-day intervals and allowed to change with sampling occasion. Since not all pic-

Table 1.5: Summary statistics of posterior distributions from spatial mark-resight model for raccoon camera trapping and telemetry data, taken from (Sollmann et al., 2013). Baseline trap encounter rate λ_0 was standardized to 7-day intervals; λ_0 and the probability of identifying a picture of a marked individual, c , were allowed to vary among the 6 sampling occasions (t); σ is estimated from telemetry data of 38 radio-collared individuals.

	Mean (SE)	2.5%	50%	97.5%
N	186.71 (14.81)	162	185	220
D	8.29 (0.66)	7.19	8.22	9.77
λ_0 ($t=1$)	0.24 (0.05)	0.16	0.23	0.34
λ_0 ($t=2$)	0.40 (0.08)	0.26	0.39	0.57
λ_0 ($t=3$)	0.11 (0.03)	0.06	0.11	0.17
λ_0 ($t=4$)	0.30 (0.07)	0.17	0.29	0.46
λ_0 ($t=5$)	0.03 (0.01)	0.02	0.03	0.06
λ_0 ($t=6$)	0.03 (0.01)	0.02	0.03	0.05
σ	0.49 (0.01)	0.47	0.49	0.51
c ($t=1$)	0.55 (0.09)	0.38	0.55	0.71
c ($t=2$)	0.39 (0.11)	0.18	0.39	0.62
c ($t=3$)	0.29 (0.11)	0.11	0.29	0.52
c ($t=4$)	0.38 (0.16)	0.10	0.36	0.71
c ($t=5$)	0.38 (0.16)	0.10	0.36	0.71
c ($t=6$)	0.30 (0.14)	0.08	0.29	0.60

tures of marked raccoons could be identified to the individual level, the authors applied the correction factor c as described in Sec. 1.4, estimated separately for each occasion.

Camera-traps recorded 117 pictures of unmarked raccoons, 33 pictures of 18 marked and identifiable raccoons, and 49 records of marked but not individually identifiable individuals (Fig. 1.3). An average of 16.32 telemetry locations (SD 4.91) were collected for each of the 38 radio-collared individuals. Raccoon abundance on the island was estimated at 186.71 (SE 14.81) individuals, which translated to a density of 8.29 (SE 0.66) individuals per km^2 . Parameter estimates are listed in Table 1.5.

In this study, although a large number of raccoons were tagged, photographic data of these tagged individuals were surprisingly sparse. Analysis of the photographic data set without the telemetry data did not render usable estimates as parallel Markov chains did not converge. One reason for the relatively sparse data was the camera trap study design: traps were spaced on average 1.77 km apart, which is about 3.5 times σ . Consequently, very few individual raccoons were photographed at more than one trap. Under these circumstances, the telemetry data provide the necessary spatial information to estimate σ and the activity centers of individual animals and thus make other model parameter estimable. Similarly, in a camera-trapping study on Florida panthers (*Puma concolor coryi*), Sollmann et al. (in revision), including telemetry data from the

3 individuals that were collared and known to use the study area resulted in density estimates with considerably higher precision as compared to preliminary estimates *without* telemetry location data, reducing the width of the 95 % BCI by about 60 %. Such improvements in precision of estimates is especially important when we are interested in changes in the population over time.

1.7 Point process models for marked individuals

As discussed in Sec. 1.1.4, all previously developed SMR models assume that marked individuals are a random sample, both spatially and demographically, from the population of the state-space. For many studies it may not be feasible to strive to meet or approximate the assumption of spatial randomness and it is thus important to generalize SMR models to situations where marking does not take place throughout \mathcal{S} . We already stated that in this situation, we generally cannot assume that activity centers of marked individuals (and unmarked, for that matter) follow a homogeneous point process. In this final section, we will describe two possible approaches to formulating such an inhomogeneous point process model. We will only provide conceptual descriptions, not a full-blown model development, as at the time of writing this book, these approaches are still somewhat experimental.

1.7.1 Homogeneous point process in a subset of \mathcal{S}

Imagine we perform an area search in a square, \mathcal{B} , for some species we want to study, maybe a reptile, and we mark all individuals we encounter. We conduct our sampling in a way that we can assume that marked individuals are randomly sampled within \mathcal{B} , and that there are no marked individuals with activity centers outside of \mathcal{B} . This design entails the assumption that \mathcal{B} can be clearly defined. We will come back to these assumptions in a minute. We then perform resighting surveys of some sort in an area that overlaps \mathcal{B} , so that, when we set a state-space around the resighting locations, \mathcal{B} is completely contained within \mathcal{S} (Fig. 1.4). We further assume that individuals that were marked in \mathcal{B} continue to live within \mathcal{B} when resighting surveys are conducted, i.e. their activity centers do not shift during the complete mark-resight study. This implies population closure across both the marking and the resighting part of the study, and in this situation we can treat the number of marked animals, m , as known. Let the total population of \mathcal{B} be N_B . Under the conditions specified above, the number of marked animals m can be described as the outcome of a binomial random variable

$$m \sim \text{Binomial}(\theta, N_B)$$

where θ is the probability that an animal living in \mathcal{B} is marked. Under these conditions we can describe the point process for marked individuals as uniform across \mathcal{B} , with zero probability of a marked activity center being located outside of \mathcal{B} . If the combined (or marginal) point process of marked and unmarked

individuals is homogeneous across \mathcal{S} , i.e. overall density is constant, then, colloquially speaking, the point process for unmarked individuals is the complement of the marked process: outside of \mathcal{B} , unmarked animals occur at the average density of \mathcal{S} , D , while inside \mathcal{B} they occur at $D * (1 - \theta)$.

The above model is an approach to specifying a spatial reference frame for marked individuals that is independent of \mathcal{S} . Some of the assumptions of the model, however, are reminiscent of traditional capture-recapture and thus, suffer from the same shortcomings. \mathcal{B} needs to be clearly defined as the area the marked individuals live in, but how do we define it? Imagine again that \mathcal{B} is a square search plot. Surely, we could capture an individual at the edge of the plot, whose activity center is located *off* that plot. Not accounting for this effect would overestimate density in \mathcal{B} . This is the equivalent of having to define an effective area sampled in traditional capture-recapture in order to estimate density. Further, we assume that θ , the probability of an individual within the plot being marked, is the same for all individuals in \mathcal{B} . But we discussed early on in this book that this is unlikely to be true, because exposure to sampling depends on an individual's home range overlap with the sampled area. So individuals near the edge of \mathcal{B} are less likely to be marked than those in the center, assuming we dispense marking effort uniformly across \mathcal{B} . In spite of the shortcomings of this approach, we believe it could serve as a reasonable approximation of some study systems. Moreover, it serves as a conceptual device because it presents a relatively simple way of thinking about two overlapping point processes, in the context of SMR the point processes describing the distribution of marked and unmarked individuals in \mathcal{S} .

1.7.2 Inhomogenous point processes

An alternative, and more realistic, point process for marked individuals is one that describes a decline in the density of marks with increasing distance to the marking location(s). We would expect this kind of spatial pattern in marks to arise, because animals living in the center of the marking grid have a higher probability to be marked than those living on the edges, which in turn have a higher marking probability than those living beyond the marking grid. As a consequence, the density of marks is higher in the center of the marking grid and decreases as we move away from it. Imagine that marking of animals takes place across some area or grid, and let C_m be the centroid of that marking area. Then, a plausible model for the distribution of marked animals is a bivariate normal model with mean C_m and variance-covariance matrix, Σ_C , where the variance in both dimensions is σ_C^2 and the covariance is 0 (Fig. 1.5). Of course, there are alternative models to describe a decrease in density, such as a negative-exponential or hazard function. Once the distribution of marked animals conditional on C_m is adequately described, the inhomogeneous point process for the unmarked animals can again be modeled by assuming that the marginal density across \mathcal{S} is constant, similar to the example above. The parameters of the inhomogeneous point process could also be modeled using the methods outlined in Chapt. ???. Density under such a model should be invariant

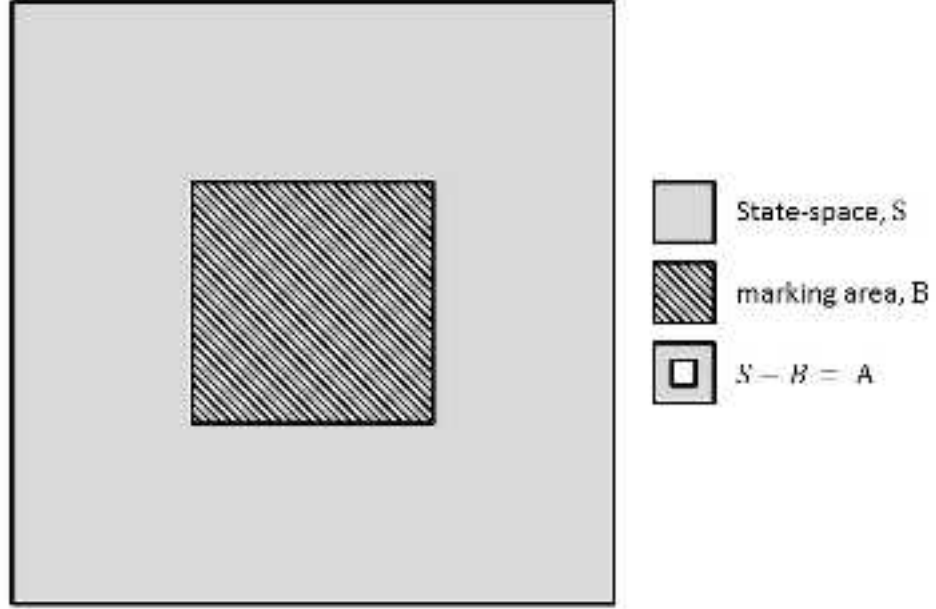


Figure 1.4: Relationship between marking area B and state-space, S .

to the size of S , as is the case with standard SCR models, because the marked individuals are probabilistically constrained to live in the vicinity of the marking locations, no matter how large we choose S .

While this model formulation is more realistic and general, it has some minor drawbacks. Based on very limited experience, the variance parameter, σ_C^2 , is barely estimable under realistic sample sizes. This is not surprising, given that this parameter attempts to describe the distribution of the activity centers, which are themselves latent. Further, the approach does entail some design constraints for marking. We implicitly assume that the centroid of the marking array is the area of highest probability of being marked, which means that marking effort has to be somewhat constant across the marking area. The assumption seems reasonable for systematic or random marking arrays (Fig. 1.5), but, for example, for a “hollow grid” the centroid would not be an adequate representation of the point of highest marking probability. The take-home message here is that an adequate point process model for the distribution of marks depends on the spatial context of the marking process.

Moving activity centers: One issue we have not addressed explicitly in this section is what to do when enough time passes between marking and resighting so that animals may have rearranged themselves spatially. In example 1 this could mean, marked animals have shifted their activity center outside of the marking plot B . In the second example, marked animals may no longer follow

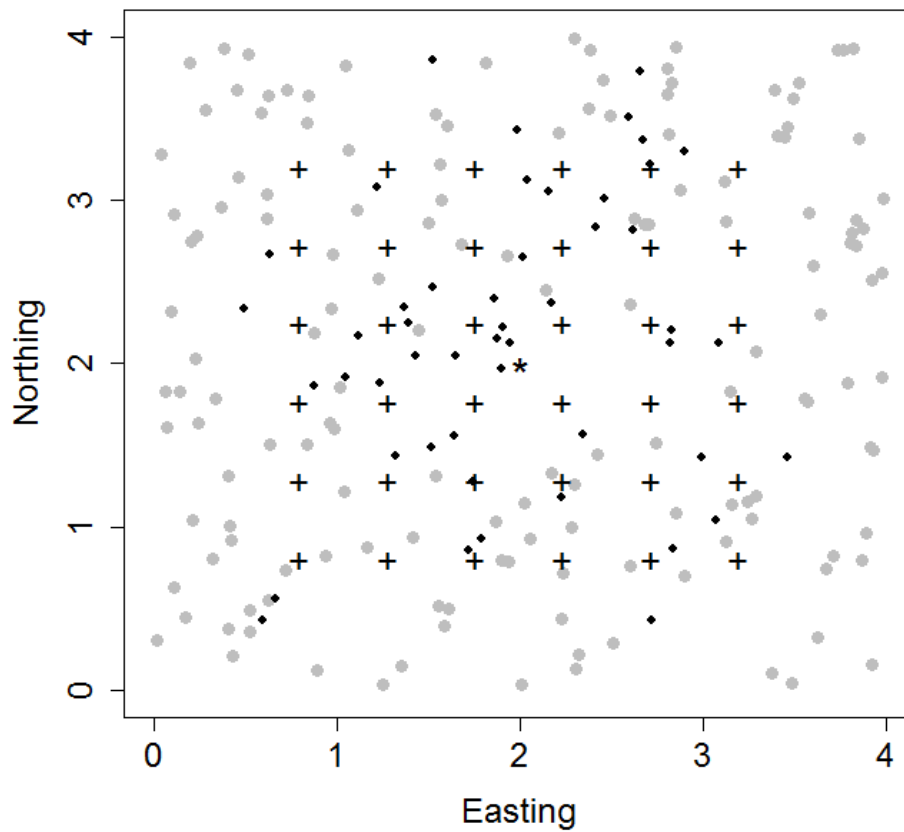


Figure 1.5: Plot of activity centers of marked (black dots) and unmarked (gray dots) individuals, with marked individuals following a bivariate normal distribution around the centroid (asterisk) of the marking locations (+); the marginal point process of both marked and unmarked individuals combined is homogeneous.

that initial distribution around the centroid of the marking array. Irrespec-
tively, we will have to make some assumption about how marked individuals
are distributed in space. We saw such a case in the Canada geese example in
Sec. 1.3.1. In that example, we defined a reasonable state-space and assumed
that animals redistributed themselves randomly across that state-space between
marking and resighting, because they were captured while molting. This seems
like a sensible solution for that particular study system, but it leads to a model
formulation where estimates of density are sensitive to choice of \mathcal{S} . Alterna-
tively, we could again use a bivariate normal distribution around the centroid
of the original marking array – if movement of activity centers is random with
respect to C_m it seems plausible that the overall underlying distribution of \mathbf{s}
would still be adequately described with a bivariate normal model. It should
be clear that choice of a model for the inhomogeneous point process of marked
individuals depends both on the spatial context of marking and what we know
(or believe) about the biology of the study species.

1.8 Summary and Outlook

In this chapter we combined SCR models (for marked individuals) with a model
for unmarked individuals, to derive a spatial mark-resight (SMR) model, which
accommodates that only part of the population is individually identifiable, usu-
ally through artificial tags. Under the assumption that marked individuals are
a random sample, both demographically and spatially, from the state-space, the
basic model with known number of marked individuals and perfect individual
identification of marks is easily modified for situations where the number of
marked individuals is unknown, or where marked animals can sometimes not
be identified to individual level. As expected, having marked individuals in
the study population improved accuracy and precision of parameter estimates
when compared to fully unmarked populations, but we also saw that the spa-
tial counts of unmarked individuals still contribute information to parameter
estimates. We further presented an approach to incorporate telemetry location
data into the spatial mark-resight model to inform estimates of σ and activity
centers. Just as in SCR, the spatial mark-resight models can account for a vari-
ety of factors that may influence individual movement and detection, as well as
survey-related parameters, and we saw one example for the Canada geese, where
 σ was sex-specific, and another for raccoons, where λ_0 was time-dependent.

Many details of SMR models remain to be explored and we noted a few of
those topics throughout this chapter. For example, we mentioned the assign-
ment of marked but unidentified records to actual marked individuals based on
their spatial location, which provides some (though imperfect) information of
their identity (Sec. 1.4). Similarly, records where the marked status cannot be
determined could potentially be included in the model as some form of overall
correction factor on detection. GPS telemetry devices and their ability to col-
lect location data with much higher frequency offer the opportunity to assign
records of collared animals to individuals based on how close to a given camera

the collared individuals were, both in space and time. In this scenario, individual identity itself could be expressed probabilistically, leading to an SMR model accounting for potential misidentification. All these possible extensions can tailor SMR models to specific survey techniques.

A fundamental assumption of the SMR models developed in this chapter was that marked animals are a random sample from \mathcal{S} . This simplifies the model as we can assume a homogeneous point process for both the marked and the unmarked part of the population. While this is a convenient situation, it is neither likely to arise often in real life, nor strictly necessary. If marked animals are not a random sample from \mathcal{S} , we need to describe their distribution in the state-space using an adequate point process that will almost always be inhomogeneous across \mathcal{S} . We mention two possible approach – a uniform (homogeneous) point process over a smaller area within \mathcal{S} , and an inhomogeneous point process where the intensity decreases as the distance to the centroid of the marking locations increases. Both formulations effectively attempt to describe the distribution of marks in space as a consequence of the spatial nature of the marking process. We believe that another way to approach this problem is to combine spatially explicit models for the marking process and the resighting process. Where marked animals were captured carries information on their spatial distribution, and it should be possible to make use of this information by formulating an integrated spatial capture-mark-resighting model. Such approaches have been developed in a non-spatial CR framework (Matechou et al., 2013; Pledger et al., 2009), but to our knowledge have not yet been addressed in the context of SCR.

Spatial mark-resight models are a recent development, and work on how to relax the spatial component of the random sample assumption and formulate adequate point process models for the distribution of marks is ongoing. While there is still a lot of work to be done, we believe that SMR modeling holds the potential to address a wide range of population estimation problems when dealing with animals that cannot be identified based on natural marks, a situation that is typical of a majority of animal population studies.

Bibliography

- 969 Arnason, A., Schwarz, C., and Gerrard, J. (1991), “Estimating closed population
970 size and number of marked animals from sighting data,” *Journal of Wildlife*
971 *Management*, 55, 716–730.
- 972 Bartmann, R. M., White, G. C., Carpenter, L. H., and Garrott, R. A. (1987),
973 “Aerial Mark-Recapture Estimates of Confined Mule Deer in Pinyon-Juniper
974 Woodland,” *The Journal of Wildlife Management*, 51, 41–46.
- 975 Bowden, D. C. (1993), “A simple technique for estimating population
976 size. Technical Report 93/12.” Tech. rep., Department of Statistics, Colorado
977 State University, Fort Collins, Colorado, USA.
- 978 Chandler, R. B. and Royle, J. A. (2013), “Spatially-explicit models for infer-
979 ence about density in unmarked or partially marked populations,” *Annals of*
980 *Applied Statistics*, xxxx–xxxx.
- 981 Hestbeck, J. B. and Malecki, R. A. (1989), “Mark-resight estimate of Canada
982 Goose midwinter numbers,” *Journal of Wildlife Management*, 53, 749–752.
- 983 Krebs, C. J. (1999), *Ecological methodology*, Menlo Park, CA: Ben-
984 jamin/Cummings.
- 985 Link, W. A., Yoshizaki, J., Bailey, L. L., and Pollock, K. H. (2010), “Uncovering
986 a latent multinomial: analysis of markrecapture data with misidentification,”
987 *Biometrics*, 66, 178–185.
- 988 Lukacs, P. M. and Burnham, K. P. (2005), “Estimating population size from
989 DNA-based closed capture-recapture data incorporating genotyping error,”
990 *Journal of Wildlife Management*, 69, 396–403.
- 991 Mace, R., Minta, S., Manley, T., and Aune, K. (1994), “Estimating grizzly bear
992 population size using camera sightings,” *Wildlife Society Bulletin*, 22, 74–83.
- 993 Matechou, E., Morgan, B. J. T., Pledger, S., Collazo, J. A., and Lyons,
994 J. E. (2013), “Integrated Analysis of Capture-Recapture-Resighting Data and
995 Counts of Unmarked Birds at Stop-Over Sites,” *Journal of Agricultural, Bi-*
996 *ological, and Environmental Statistics*.

- 997 McClintock, B. and Hoeting, J. (2010), “Bayesian analysis of abundance for
998 binomial sighting data with unknown number of marked individuals,” *Envi-
999 ronmental and Ecological Statistics*, 17, 317–332.
- 1000 McClintock, B. and White, G. (2012), “From NOREMARK to MARK: soft-
1001 ware for estimating demographic parameters using mark-resight methodol-
1002 ogy,” *Journal of Ornithology*, 152, 641–650.
- 1003 McClintock, B., White, G., Antolin, M., and Tripp, D. (2009a), “Estimating
1004 abundance using mark-resight when sampling is with replacement or the num-
1005 ber of marked individuals is unknown,” *Biometrics*, 65, 237–246.
- 1006 McClintock, B., White, G., and Burnham, K. (2006), “A robust design mark-
1007 resight abundance estimator allowing heterogeneity in resighting probabili-
1008 ties,” *Journal of agricultural, biological, and environmental statistics*, 11,
1009 231–248.
- 1010 McClintock, B. T., White, G. C., Burnham, K. P., and Pryde, M. A. (2009b),
1011 “A generalized mixed effects model of abundance for mark-resight data when
1012 sampling is without replacement,” in *Modeling demographic processes in
1013 marked populations*, ed. Thomson, D., New York: Springer, pp. 271–289.
- 1014 Minta, S. and Mangel, M. (1989), “A Simple Population Estimate Based on
1015 Simulation for Capture-Recapture and Capture-Resight Data,” *Ecology*, 70,
1016 1738–1751.
- 1017 Neal, A., White, G., Gill, R., Reed, D., and Olterman, J. (1993), “Evaluation
1018 of mark-resight model assumptions for estimating mountain sheep numbers,”
1019 *Journal of Wildlife Management*, 57, 436–450.
- 1020 Neal, A. K. (1990), “Evaluation of mark-resight population estimates using
1021 simulations and field data from mountain sheep.” M.S. thesis, Colorado State
1022 University, Fort Collins, Colorado, USA.
- 1023 Pledger, S., Efford, M., Pollock, K., Collazo, J., and Lyons, J. (2009), “Stopover
1024 duration analysis with departure probability dependent on unknown time
1025 since arrival,” *Modeling demographic processes in marked populations*, 3, 349–
1026 363.
- 1027 Rutledge, M. (2013), “Impacts of Resident Canada Geese in a Suburban Envi-
1028 ronment,” Ph.D. thesis, North Carolina State University.
- 1029 Skalski, J. R., Millspaugh, J. J., and Spencer, R. D. (2005), “Population estima-
1030 tion and biases in paintball, mark-resight surveys of elk,” *Journal of Wildlife
1031 Management*, 69, 1043–1052.
- 1032 Sollmann, R., Gardner, B., Chandler, R. B., Shindle, D., Onorato, D. P., Royle,
1033 J. A., and O’Connell, A. F. (in revision), “Using multiple data sources pro-
1034 vides density estimates for endangered Florida panther,” *Journal of Applied
1035 Ecology*.

- 1036 Sollmann, R., Gardner, B., Parsons, A., Stocking, J., McClintock, B., Simons,
1037 T., Pollock, K., and O'Connell, A. (2013), "A spatial mark-resight model
1038 augmented with telemetry data," *Ecology*.
- 1039 White, G. (1996), "NOREMARK: population estimation from mark-resighting
1040 surveys," *Wildlife Society Bulletin*, 24, 50–52.
- 1041 White, G. and Shenk, M. (2001), "Population estimation with radio-marked
1042 individuals." in *Radio tracking and animal populations*, eds. Millspaugh, J.
1043 and Marzluff, J., San Diego, USA: Academic Press, pp. 329–350.
- 1044 White, G. C. and Garrot, R. (1990), *Analysis of wildlife radiolocation data*, New
1045 York, USA: Academic Press.
- 1046 Yoshizaki, J., Pollock, K. H., Brownie, C., and Webster, R. A. (2009), "Mod-
1047 eling misidentification errors in capture-recapture studies using photographic
1048 identification of evolving marks," *Ecology*, 90, 3–9.