

# <sup>1</sup> Chapter 1

## <sup>2</sup> Introduction



## <sup>3</sup> Chapter 2

## <sup>4</sup> GLMS and WinBUGS



## Chapter 3

# Closed Population Models

Having covered the basics of hierarchical models and their implementation, in this chapter we will consider ordinary capture-recapture (CR) models for estimating population size in closed populations. We will see that such models are closely related to binomial (or logistic) regression type models. In fact, when  $N$  is known, they are precisely such models. We consider some important extensions of ordinary closed population models that accommodate various types of “individual effects” — either in the form of explicit covariates (sex, age, body mass) or unstructured “heterogeneity” in the form of an individual random effect. In general, these models are variations of generalized linear or generalized linear mixed models (GLMMs). Because of the paramount importance of this concept, we focus mainly on fairly simple models in which the observations are individual encounter frequencies,  $y_i$  = the number of encounters of individual  $i$  out of  $K$  replicate samples of the population which, for the models we consider here, is the outcome of a binomial random variable. Along the way, we consider the spatial context of capture-recapture data and models and demonstrate that density cannot be formally estimated when spatial information is ignored. We also review some of the informal methods of estimating density using CR methods, and consider some of their limitations. We will be exposed to our first primitive spatial capture-recapture models which arise as relatively minor variations of so-called “individual covariate models” (of the Huggins (1989) and Alho (1990) variety). In a sense, the point of this chapter is to establish that linkage between non-spatial and spatial capture-recapture models in a direct and concise manner beginning with the basic “model  $M_0$ ” and extensions of that model to include individual heterogeneity and also individual covariates. A special type of individual covariate models is distance sampling, which could be thought of as the most primitive spatial capture-recapture model. In later chapters we further develop and extend ideas introduced in this chapter.

We emphasize Bayesian analysis of capture-recapture models and we accomplish this using a method related to classical “data augmentation” from the

statistics literature (e.g., Tanner and Wong, 1987)). This is a general concept in statistics but, in the context of capture-recapture models where  $N$  is unknown, it has a consistent implementation across classes of capture-recapture models and one that is really convenient from the standpoint of doing MCMC (Royle et al., 2007). We use data augmentation throughout this book and thus emphasize its conceptual and technical origins and demonstrate applications to closed population models. We refer the reader to Kéry and Schaub (2011, ch. 6) for an accessible and complimentary development of ordinary closed population models.

### 3.1 The Simplest Closed Population Model: Model $M_0$

To start looking at the simplest capture-recapture model, let's suppose there exists a population of  $N$  individuals which we subject to repeated sampling, say over  $K$  nights, where individuals are captured, marked, and subsequently recaptured. We suppose that individual encounter histories are obtained, and these are of the form of a sequence of 0's and 1's indicating capture ( $y = 1$ ) or not ( $y = 0$ ) during any sampling occasion ("sample"). As an example, suppose  $K = 5$  sampling occasions, then an individual captured during sample 2 and 3 but not otherwise would have an encounter history of the form  $\mathbf{y} = (0, 1, 1, 0, 0)$ . Thus, the observation  $\mathbf{y}_i$  for each individual ( $i$ ) is a vector having elements denoted by  $y_{ik}$  for  $k = 1, 2, \dots, K$ . Usually this is organized as a row of a matrix with elements  $y_{ik}$ , see Table 3.1. Except where noted explicitly, we suppose that observations are independent within individuals and among individuals. Formally, this allows us to say that  $y_{ik}$  are *iid* Bernoulli random variables and we may write  $y_{ik} \sim \text{Bern}(p)$ . Consequently, for this very simple model in which  $p$  is in fact constant, then we can declare that the individual encounter frequencies (total captures),  $y_i = \sum_k y_{ik}$ , have a binomial distribution based on a sample of size  $K$ . That is

$$y_i = \sum_k y_{ik} \sim \text{Bin}(p, K)$$

for every individual in the population. This is a remarkably simple model that forms the cornerstone of almost all of classical capture-recapture models, including most spatial capture-recapture models discussed throughout this book.

Evidently, the basic capture-recapture model structure is precisely a simplistic version of a logistic-regression model with only an intercept term (logit( $p$ ) = constant). To say that all capture-recapture models are just logistic regressions is only slightly inaccurate. In fact, we are proceeding here "conditional on  $N$ ", i.e., as if we knew  $N$ . In practice we don't, of course, and that is kind of the point of capture-recapture models as estimating  $N$  is the central objective. But, by proceeding conditional on  $N$ , we can specify a simple model and then deal with the fact that  $N$  is unknown using standard methods that you are already familiar with (i.e., GLMs - see Chapt. 2).

Table 3.1: a capture-recapture data set with  $n = 6$  observed individuals and  $K = 5$  samples.

| indiv $i$ | Sample occasion |   |   |   |   | $y_i$ |
|-----------|-----------------|---|---|---|---|-------|
|           | 1               | 2 | 3 | 4 | 5 |       |
| 1         | 1               | 0 | 0 | 1 | 0 | 2     |
| 2         | 0               | 1 | 0 | 0 | 1 | 2     |
| 3         | 1               | 0 | 0 | 1 | 0 | 2     |
| 4         | 1               | 0 | 1 | 0 | 1 | 3     |
| 5         | 0               | 1 | 0 | 0 | 0 | 1     |
| $n = 6$   | 1               | 0 | 0 | 0 | 0 | 1     |

Assuming individuals of the population are observed independently, the joint probability distribution of the observations is the product of  $N$  binomials

$$\begin{aligned} \Pr(y_1, \dots, y_N | p) &= \prod_{i=1}^N \text{Bin}(y_i | K, p) \\ &= \prod_{k=0}^K \pi(k)^{n_k} \end{aligned}$$

where  $\pi(k) = \text{Bin}(k | K, p)$  and where  $n_k = \sum_{i=1}^N I(y_i = k)$  denotes the number of individuals captured  $k$  times in  $K$  surveys. We emphasize that this is conditional on  $N$ , in which case we get to observe the  $y = 0$  observations and the resulting data are just *iid* binomial counts. Because this is a binomial regression model of the variety described in Chapt. ??, fitting this model using a **BUGS** engine poses no difficulty.

The essential problem in capture-recapture, however, is that  $N$  is not known because the number of uncaptured/missing individuals (i.e., those in the zero cell that occur with probability  $\pi(0)$ ) is unknown. Consequently, the observed capture frequencies  $n_k$  are no longer independent. Instead, their joint distribution is multinomial (e.g., see Illian et al. (2008, p. xyz)):

$$n_1, n_2, \dots, n_K \sim \text{Multin}(N, \pi(1), \pi(2), \dots, \pi(K)) \quad (3.1)$$

We denote the number of uncaptured/missing individuals by  $n_0$ , and the total number of distinct individuals encountered in the  $K$  samples by  $n = \sum_{k=1}^K n_k$ . Note that  $n_0$  appears in the likelihood as a component of  $N = n + n_0$ .

To fit the model in which  $N$  is *unknown*, we can regard  $n_0$  as a parameter and maximize the multinomial likelihood directly. While direct likelihood analysis of the multinomial model is straightforward, that does not prove to be too useful in practice because we seldom are concerned with models for the aggregated encounter history frequencies, which entail that capture probabilities are the same for all individuals. In many instances, including for spatial capture-recapture (SCR) models, we require a formulation of the model that can accommodate individual level covariates to account for differences in detection among individuals which we address subsequently in this chapter.

Table 3.2: Modes of analysis of capture-recapture models. Closed population models can be analyzed using the joint or “full likelihood” which contains  $N$  as an explicit parameter, the conditional likelihood which does not involve  $N$ , or by data augmentation which replaces  $N$  with  $\psi$ . Each approach yields a distinct likelihood.

| Mode of analysis       | parameters in model | statistical model          |
|------------------------|---------------------|----------------------------|
| Joint likelihood       | $p, N$              | multinomial with index $N$ |
| Conditional likelihood | $p$                 | zero-truncated binomial    |
| Data augmentation      | $p, \psi$           | zero-inflated binomial     |

### 3.1.1 Conditional likelihood

We saw that a basic closed population model is a simple logistic regression model if  $N$  is known and, when  $N$  is unknown, the model is multinomial with index or sample size parameter  $N$ . This multinomial model, being conditional on  $N$ , is sometimes referred to as the “joint likelihood” the “full likelihood” or the “unconditional likelihood” (or model in place of likelihood). This formulation differs from the so-called “conditional likelihood” approach in which the likelihood of the observed encounter histories is devised conditional on the event that an individual is captured at least once. To construct this likelihood, we have to recognize that individuals appear or not in the sample based on the value of the random variable  $y_i$ , that is, we capture them if and only if  $y_i > 0$ . The observation model is therefore based on  $\Pr(y|y > 0)$ . For the simple case of model  $M_0$ , the resulting conditional distribution is a “zero truncated” binomial distribution which accounts for the fact that we cannot observe the value  $y = 0$  in the data set (see Royle and Dorazio, 2008, sec. 5.1). Both the conditional and unconditional models are legitimate modes of analysis in all capture-recapture types of studies, and they provide equally valid descriptions of the data and for many practical purposes provide equivalent inferences, at least in large sample sizes (Sanathanan, 1972).

In this book we emphasize Bayesian analysis of capture-recapture models using data augmentation (discussed subsequently), which produces yet a third distinct formulation of capture recapture-models based on the zero-*inflated* binomial distribution that we describe in the next section. Thus, there are 3 distinct formulations of the model – or modes of analysis – for analyzing all capture-recapture models based on the (1) binomial model for the joint or unconditional specification; (2) zero-truncated binomial that arises “conditional on  $n$ ”; and (3) the zero-inflated binomial that arises under data augmentation. Each formulation has a distinct complement of model parameters (shown in Table 3.2 for model  $M_0$ ).



## 3.2 Data Augmentation

We consider a method of analyzing closed population models using data augmentation (DA) which is useful for Bayesian analysis and, in particular, analysis of models using the various **BUGS** engines and other software. Data augmentation is a general statistical concept that is widely used in statistics in many different settings. The classical reference is Tanner and Wong (1987) but see also Liu and Wu (1999). Data augmentation can be adapted to provide a very generic framework for Bayesian analysis of capture-recapture models with unknown  $N$ . This idea was introduced for closed populations by Royle et al. (2007), and has subsequently been applied to a number of different contexts including individual covariate models (Royle, 2009), open population models (Royle and Dorazio, 2008, 2010; Gardner et al., 2010a), spatial capture-recapture models (Royle and Young, 2008; Royle, 2010; Gardner, 2009), and many others. Kéry and Schaub (2011, Chapt. 6) provides a good introduction to data augmentation in the context of closed population models.

Conceptually, data augmentation is a reparameterization of the “complete data” model – that which is conditional on  $N$ . The reparameterization is achieved by embedding this data set into a larger data set having  $M > N$  “rows” (individuals) and reexpressing the model conditional on  $M$  instead of  $N$ . The great thing about data augmentation is that we do not need to know  $N$  for this reparameterization. Although this has a whiff of arbitrariness or even ad hocery to it in the choice of  $M$ , it is always possible, in practice, to choose  $M$  pretty easily for a given problem and context and results will be insensitive to choice of  $M$ <sup>1</sup>. Then, under data augmentation, analysis is focused on the “augmented data set.” That is, we analyze the bigger data set - the one having  $M$  rows - with an appropriate model that accounts for the augmentation. Inference is focused directly on estimating the proportion  $\psi = E[N]/M$ , instead of directly on  $N$ , where  $\psi$  is the “data augmentation parameter.”

### 3.2.1 DA links occupancy models and closed population models

There is a close correspondence between so-called “occupancy” models and closed population models following Royle and Dorazio (2008, sec. 5.6).

In occupancy models (MacKenzie et al., 2002; Tyre et al., 2003) the sampling situation is that  $M$  sites, or patches, are sampled multiple times to assess whether a species occurs at each site. This yields encounter data such as that illustrated in the left panel of Table 3.3. The important problem is that a species may occur at a site, but go undetected, yielding the “all-zero” encounter histories which are observed. However, some of the all-zeros may well correspond to sites where the species in fact *does* occur. Thus, while the zeros are observed, there are too many of them and, in a sense, the inference problem is to allocate the zeros into “structural” (fixed) and “sampling” (or stochastic) zeros. More

<sup>1</sup>Unless the data set is sufficiently small that parameters are weakly identified

formally, inference is focused on the parameter  $\psi$ , the probability that a site is occupied. In contrast, in classical closed population studies, we observe a data set as in the middle panel of Table 3.3 where *no* zeros are observed. The inference problem is, essentially, to estimate how many sampling zeros there are - or should be - in a “complete” data set. This objective (how many sampling zeros?) is precisely the same for both types of problems if an upper limit  $M$  is specified for the closed population model. The only distinction being that, in occupancy models,  $M$  is set by design (i.e., the number of sites to visit) whereas a natural choice of  $M$  for capture-recapture models may not be obvious. However, by assuming a uniform prior for  $N$  on the integers  $[0, M]$ , this upper bound is induced (Royle et al., 2007). Then, one can analyze capture-recapture models by adding  $M - n$  all-zero encounter histories to the data set and regarding the augmented data set, essentially, as a site-occupancy data set.

Thus, the heuristic motivation of data augmentation is to fix the size of the data set by adding *too many* all-zero encounter histories to create the data set shown in the right panel of Table 3.3 - and then analyze the augmented data set using an occupancy type model which includes both “unoccupied sites” as well as “occupied sites” at which detections did not occur. We call these  $M - n$  all-zero histories “potential individuals” because they exist to be recruited (in a non-biological sense) into the population, for example during an analysis by MCMC.

To analyze the augmented data set, we recognize that it is a zero-inflated version of the known- $N$  data set. That is, some of the augmented all-zeros are sampling zeros (corresponding to actual individuals that were missed) and some are “structural” zeros, which do not correspond to individuals in the population. For a basic closed-population model, the resulting likelihood under data augmentation - that is, for the data set of size  $M$  - is a simple zero-inflated binomial likelihood. The zero-inflated binomial model can be described “hierarchically”, by introducing a set of binary latent variables,  $z_1, z_2, \dots, z_M$ , to indicate whether each individual  $i$  is ( $z_i = 1$ ) or is not ( $z_i = 0$ ) a member of the population of  $N$  individuals exposed to sampling. We assume that  $z_i \sim \text{Bern}(\psi)$  where  $\psi$  is the probability that an individual in the data set of size  $M$  is a member of the sampled population - in the sense that  $1 - \psi$  is the probability of realizing a “structural zero” in the augmented data set. The zero-inflated binomial model which arises under data augmentation can be formally expressed by the following set of assumptions:

$$\begin{aligned} y_i | z_i = 1 &\sim \text{Bin}(K, p) \\ y_i | z_i = 0 &\sim \delta(0) \\ z_i &\stackrel{iid}{\sim} \text{Bern}(\psi) \\ \psi &\sim \text{Unif}(0, 1) \\ p &\sim \text{Unif}(0, 1) \end{aligned}$$

for  $i = 1, \dots, M$ , where  $\delta(0)$  is a point mass at  $y = 0$ .

208 Note that, under data augmentation,  $N$  is no longer an explicit parameter  
 209 of this model. Instead, we estimate  $\psi$  and functions of the latent variables. In  
 210 particular, under the assumptions of the zero-inflated model,  $z_i \stackrel{iid}{\sim} \text{Bern}(\psi)$ ;  
 211 therefore,  $N$  is a function of these latent variables:

$$N = \sum_{i=1}^M z_i.$$

212 Further, we note that the latent  $z_i$  parameters *can be* removed from the model  
 213 by integration, in which case the joint probability of the data is

$$\Pr(y_1, \dots, y_M | p, \psi) = \prod_{i=1}^M \psi \text{Bin}(y_i | K, p) + I(y_i = 0)(1 - \psi) \quad (3.2)$$

214 Which can be maximized directly to obtain the MLEs of the structural param-  
 215 eters  $\psi$  and  $p$  or those of other more complex models (e.g., see Royle, 2006). We  
 216 could estimate these parameters and then use them to obtain an estimator of  
 217  $N$  using the so-called “Best unbiased predictor” (see Royle and Dorazio, 2011).  
 218 Normally, however, we will analyze the model in its “conditional-on- $z$ ” form  
 219 using methods of MCMC either in the **BUGS** engines or using our own MCMC  
 220 algorithms (see Chapt. 7).

### 221 3.2.2 Model $M_0$ in BUGS

222 For model  $M_0$  in which we can aggregate the encounter data to individual-  
 223 specific encounter frequencies, the augmented data are given by the vector of fre-  
 224 quencies  $(y_1, \dots, y_n, 0, 0, \dots, 0)$ . The zero-inflated model of the augmented data  
 225 combines the model of the latent variables,  $z_i \sim \text{Bern}(\psi)$  with the conditional-  
 226 on- $z$  binomial model:

$$\begin{aligned} y_i | z_i = 1 &\sim \text{Bin}(K, p) \\ y_i | z_i = 0 &\sim \delta(0) \end{aligned}$$

227 It is convenient to express the conditional-on- $z$  observation model concisely as:

$$y_i | z_i \sim \text{Bin}(K, pz_i)$$

228 Thus, if  $z_i = 0$  then the success probability of the binomial distribution is  
 229 identically 0 whereas, if  $z_i = 1$ , then the success probability is  $p$ . This is useful  
 230 in describing the model in the **BUGS** language, as shown in Panel 3.1. Note  
 231 the last line of the model specification provides the expression for computing  $N$   
 232 from the data augmentation variables  $z_i$ .

233 Specification of a more general model in terms of the individual encounter  
 234 observations  $y_{ik}$  is not much more difficult than for the individual encounter  
 235 frequencies. We define the observation model by a double loop and change the  
 236 indexing of things accordingly, i.e.,

Table 3.3: Hypothetical occupancy data set (left), capture-recapture data in standard form (center), and capture-recapture data augmented with all-zero capture histories (right).

| Occupancy data |     |     |     | Capture-recapture |     |     |     | Augmented C-R |     |     |     |
|----------------|-----|-----|-----|-------------------|-----|-----|-----|---------------|-----|-----|-----|
| site           | k=1 | k=2 | k=3 | ind               | k=1 | k=2 | k=3 | ind           | k=1 | k=2 | k=3 |
| 1              | 0   | 1   | 0   | 1                 | 0   | 1   | 0   | 1             | 0   | 1   | 0   |
| 2              | 1   | 0   | 1   | 2                 | 1   | 0   | 1   | 2             | 1   | 0   | 1   |
| 3              | 0   | 1   | 0   | .                 | 0   | 1   | 0   | 3             | 1   | 0   | 1   |
| 4              | 1   | 0   | 1   | .                 | 1   | 0   | 1   | 4             | 1   | 0   | 1   |
| 5              | 0   | 1   | 1   | .                 | 0   | 1   | 1   | 5             | 1   | 0   | 1   |
| .              | 0   | 1   | 1   | .                 | 0   | 1   | 1   | .             | 0   | 1   | 1   |
| .              | 1   | 1   | 1   | .                 | 1   | 1   | 1   | .             | 0   | 1   | 1   |
| .              | 1   | 1   | 1   | .                 | 1   | 1   | 1   | .             | 1   | 1   | 1   |
| .              | 1   | 1   | 1   | .                 | 1   | 1   | 1   | .             | 1   | 1   | 1   |
| n              | 1   | 1   | 1   | n                 | 1   | 1   | 1   | n             | 1   | 1   | 1   |
| .              | 0   | 0   | 0   |                   |     |     |     | .             | 0   | 0   | 0   |
| .              | 0   | 0   | 0   |                   |     |     |     | .             | 0   | 0   | 0   |
|                | 0   | 0   | 0   |                   |     |     |     |               | 0   | 0   | 0   |
|                | 0   | 0   | 0   |                   |     |     |     |               | 0   | 0   | 0   |
|                | 0   | 0   | 0   |                   |     |     |     |               | 0   | 0   | 0   |
|                | 0   | 0   | 0   |                   |     |     |     | N             | 0   | 0   | 0   |
| .              | 0   | 0   | 0   |                   |     |     |     | .             | 0   | 0   | 0   |
| .              | 0   | 0   | 0   |                   |     |     |     | .             | 0   | 0   | 0   |
| M              | 0   | 0   | 0   |                   |     |     |     | .             | 0   | 0   | 0   |
|                |     |     |     |                   |     |     |     | .             | .   | .   | .   |
|                |     |     |     |                   |     |     |     | .             | .   | .   | .   |
|                |     |     |     |                   |     |     |     | .             | .   | .   | .   |
|                |     |     |     |                   |     |     |     | M             | 0   | 0   | 0   |

---

```

model{
  p ~ dunif(0,1)
  psi~dunif(0,1)

  # nind = number of individuals captured at least once
  # nz = number of uncaptured individuals added for PX-DA
  for(i in 1:(nind+nz)) {
    z[i]~dbern(psi)
    mu[i]<-z[i]*p
    y[i]~dbin(mu[i],K)
  }

  N<-sum(z[1:(nind+nz)])
}

```

---

Panel 3.1: Model  $M_0$  under data augmentation.

```

237 for(i in 1:(nind+nz)) {
238   z[i]~dbern(psi)
239   for(k in 1:K){
240     mu[i,k]<-z[i]*p
241     y[i,k]~dbin(mu[i,k],1)
242   }
243 }

```

244 In this manner, it is straightforward to incorporate covariates on  $p$  for both  
 245 individuals and sampling occasions (see discussion of this below and also Chapt.  
 246 9 as well as to devise other extensions of the model.

### 247 3.2.3 Formal development of data augmentation

248 Use of DA for solving inference problems with unknown  $N$  can be justified as  
 249 originating from the choice of uniform prior on  $N$ . The  $\text{Unif}(0, M)$  prior for  $N$   
 250 is innocuous in the sense that the posterior associated with this prior is equal  
 251 to the likelihood for sufficiently large  $M$ . One way of inducing the  $\text{Unif}(0, M)$   
 252 prior on  $N$  is by assuming the following hierarchical prior:

$$\begin{aligned}
 N &\sim \text{Bin}(M, \psi) \\
 \psi &\sim \text{Unif}(0, 1)
 \end{aligned}
 \tag{3.3}$$

253 which includes a new model parameter  $\psi$  (note that we have seen  $\psi$  in the  
 254 previous section as the proportion  $E[N]/M$ ). This parameter denotes the prob-  
 255 ability that an individual in the super-population of size  $M$  is a member of  
 256 the population of  $N$  individuals exposed to sampling. The model assumptions,

specifically the multinomial model (Eq. 3.1) and Eq. 3.3, may be combined to yield a reparameterization of the conventional model that is appropriate for the augmented data set of known size  $M$ :

$$(n_1, n_2, \dots, n_K) \sim \text{Multin}(M, \psi\pi(1), \psi\pi(2), \dots, \psi\pi(K)) \quad (3.4)$$

This arises by removing  $N$  from Eq. 3.1 by integrating over the binomial prior distribution for  $N$ . Thus, the models we analyze under data augmentation arise formally by removing the parameter  $N$  from the ordinary model - the model conditional on  $N$  - by integrating over a binomial prior distribution for  $N$ .

Note that the  $M - n$  unobserved individuals in the augmented data set have probability  $\psi\pi(0) + (1 - \psi)$ , indicating that these unobserved individuals are a mixture of individuals that are sampling zeros ( $\psi\pi_0$ , and belong to the population of size  $N$ ) and others that are “structural zeros” (occurring in the augmented data set with probability  $1 - \psi$ ). In Eq. 3.4  $N$  has been eliminated as a formal parameter of the model by marginalization (integration) and replaced with the new parameter  $\psi$ , the data augmentation parameter. However, the full likelihood containing both  $N$  and  $\psi$  can also be analyzed (see Royle et al., 2007).

### 3.2.4 Remarks on Data Augmentation

Data augmentation may seem like a strange and mysterious black-box, and likely it is unfamiliar to most people, even those with substantial experience with capture-recapture models. However, it really is just a formal reparameterization of capture-recapture models in which  $N$  is removed from the ordinary (conditional-on- $N$ ) model by integration. In the case of model  $M_0$ , data augmentation produces the zero-inflated binomial which is distinct from the original observation model, but only in the sense that it embodies, explicitly, the  $\text{Unif}(0, M)$  prior for  $N$ . Choice of  $M$  might be cause for some concern related to potential sensitivity to choice of  $M$ . The guiding principle is that it should be chosen large enough so that the posterior for  $N$  is not truncated, but no larger because large values entail more computational burden. It seems likely that the properties of the Markov chains should be affected by  $M$  and so some optimality might exist (Gopalaswamy et al., 2012), as in occupancy models (Mackenzie and Royle, 2005). Formal analysis of this is needed.

We emphasize the motivation for data augmentation being that it produces a data set of fixed size, so that the parameter dimension in any capture-recapture model is also fixed. As a result, MCMC is a relatively simple proposition using standard Gibbs Sampling. Consider the simplest context - analyzing model  $M_0$  using the occupancy type model. In this case, DA converts model  $M_0$  to a basic occupancy model and the parameters  $p$  and  $\psi$  have known full-conditional distributions (in fact, beta distributions) that can be sampled from directly. Furthermore, the data augmentation variables - i.e., the collection of  $z$ 's, can be sampled from Bernoulli full conditionals. MCMC is not too much more difficult for complicated models - sometimes the hyperparameters need to be

sampled using a Metropolis-Hastings step (e.g., Chapt. 7), but nothing more sophisticated than that is required.

There are other approaches to analyzing models with unknown  $N$ , using reversible jump MCMC (RJMCMC) or other so-called “trans-dimensional” (TD) algorithms (Durban and Elston, 2005; King and Brooks, 2001; King et al., 2008; Schofield and Barker, 2008; Wright et al., 2009). What distinguishes DA from RJMCMC and related TD methods is that DA is used to create a distinctly new model that is unconditional on  $N$  and we (usually) analyze the unconditional model. The various TD/RJMCMC approaches seek to analyze the conditional-on- $N$  model in which the dimensional of the parameter space is a variable function of  $N$ . TD/RJMCMC approaches might appear to have the advantage that one can model  $N$  explicitly or consider alternative priors for  $N$ . However, despite that  $N$  is removed as an explicit parameter in DA, it is possible to develop hierarchical models that involve structure on  $N$  (Converse and Royle, 2010; Royle et al., 2011a) which we consider in Chapt. ??.

### 3.2.5 Example: Black Bear Study on Fort Drum

To illustrate the analysis of Model  $m_0$  using data augmentation, we use a data set collected at Fort Drum Military Installation in upstate New York by the Department of Defense, Cornell University and colleagues. These data have been analyzed in various forms by Wegan (2008); Gardner (2009) and Gardner et al. (2010b). The specific data used here are encounter histories on 47 individuals obtained from an array of 38 baited “hair snares” (Fig. 3.1) during June and July 2006. Barbed wire traps were baited and checked for hair samples each week for eight weeks, thus we have  $K = 8$  sample intervals. The data are provided in the **R** package `scrbook` and the analysis can be set up and run as follows. Here, the data were augmented with  $M - n = 128$  ( $M = 175$ ) all-zero encounter histories.

```
library("scrbook")
data("beardata")
trapmat<-beardata$trapmat
nind<-dim(beardata$bearArray)[1]
K<-dim(beardata$bearArray)[3]
ntraps<-dim(beardata$bearArray)[2]

M=175
nz<-M-nind
Yaug <- array(0, dim=c(M,ntraps,K))

Yaug[1:nind,,]<-beardata$bearArray
y<- apply(Yaug,c(1,3),sum) # summarize by ind x rep
y[y>1]<- 1 # toss out duplicate obs
ytot<-apply(y,1,sum) # total encounters out of K
```

The raw data object, `beardata$bearArray` is a 3-dimensional array  $nind \times ntraps \times K$  of individual encounter events (i.e.,  $y_{ijk} = 1$  if individual  $i$  was

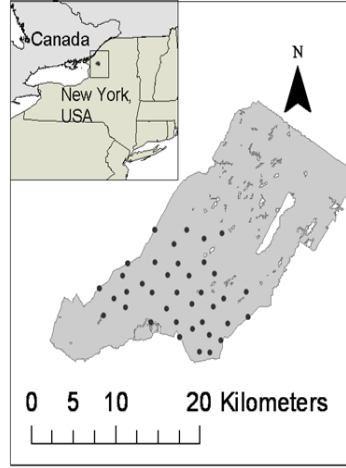


Figure 3.1: Fort Drum study area and hair snare locations.

encountered in trap  $j$  during occasion  $k$ , and 0 otherwise). For fitting model  $M_0$  (or  $M_h$ , see below), it is sufficient to reduce the data to individual encounter frequencies which we have labeled  $y_{tot}$  above. The **BUGS** model file along with commands to fit the model are as follows:

```

346 set.seed(2013)                # to obtain the same results each time
347 library("R2WinBUGS")
348 data0<-list(y=y,M=M,K=K)
349 params0<-list('psi','p','N')
350 zst=c(rep(1,nind),rbinom(M-nind, 1, .5))
351 inits = function() {list(z=zst, psi=runif(1), p=runif(1)) }
352
353 cat("
354 model {
355
356   psi~dunif(0, 1)
357   p~dunif(0,1)
358
359   for (i in 1:M){
360     z[i]~dbern(psi)
361     for(k in 1:K){
362       tmp[i,k]<-p*z[i]
363       y[i,k]~dbin(tmp[i,k],1)
364     }
365   }
366   N<-sum(z[1:M])
367 }
368 ",file="modelM0.txt")
369
```



```

370 fit0 = bugs(data0, inits, params0, model.file="modelM0.txt",
371             n.chains=3, n.iter=2000, n.burnin=1000, n.thin=1,
372             debug=TRUE, working.directory=getwd())

```

373 This produces the follow posterior summary statistics:

```

374 > print(fit0,digits=2)
375 Inference for Bugs model at "modelM0.txt", fit using WinBUGS,
376 3 chains, each with 2000 iterations (first 1000 discarded)
377 n.sims = 3000 iterations saved
378      mean    sd  2.5%   25%   50%   75%  97.5% Rhat n.eff
379 psi      0.29 0.04  0.22  0.26  0.29  0.31  0.36   1  3000
380 p        0.30 0.03  0.25  0.28  0.30  0.32  0.35   1  3000
381 N        49.94 1.99 47.00 48.00 50.00 51.00 54.00   1  3000
382 deviance 489.05 11.28 471.00 480.45 488.80 495.40 513.70   1  3000
383
384 [... some output deleted ...]

```

385 **WinBUGS** did well in choosing an MCMC algorithm for this model – we  
386 have  $\hat{R} = 1$  for each parameter, and an effective sample size of 3000, equal to  
387 the total number of posterior samples<sup>2</sup>. We see that the posterior mean of  $N$   
388 under this model is 49.94 and a 95% posterior interval is (48, 54). We revisit  
389 these data later in the context of more complex models.

390 In order to obtain an estimate of density,  $D$ , we need an area to associate  
391 with the estimate of  $N$ , and in Chapt. 1 we already went through a number of  
392 commonly used procedures to conjure up such an area, including buffering the  
393 trap array by the home range radius, often estimated by the mean maximum  
394 distance moved (MMDM) (Parmenter et al., 2003), 1/2 MMDM (Dice, 1938)  
395 or directly from telemetry data (Wallace et al., 2003) Typically, the trap array  
396 is defined by the convex hull around the trap locations, and this is what we  
397 applied a buffer to. We computed the buffer by using an estimate of the mean  
398 female home range radius (2.19 km) estimated from telemetry studies (Bales  
399 et al., 2005) instead of using an estimate based on our relatively more sparse  
400 recapture data. For the Fort Drum study, the convex hull has area 157.135  
401  $km^2$ , and the buffered convex hull has area 277.011  $km^2$ . To create this we  
402 used functions contained in the **R** package **rgeos** and created a utility function  
403 **bcharea** which is in our **R** package **scrbook**. The commands are as follows:

```

404 library("rgeos")
405
406 bcharea<-function(buff,traplocs){
407   p1<-Polygon(rbind(traplocs,traplocs[1,]))
408   p2<-Polygons(list(p1=p1),ID=1)
409   p3<-SpatialPolygons(list(p2=p2))
410   p1ch<-gConvexHull(p3)
411   bp1<-gBuffer(p1ch, width=buff)
412   plot(bp1, col='gray')

```

---

<sup>2</sup>This is even a little suspicious....

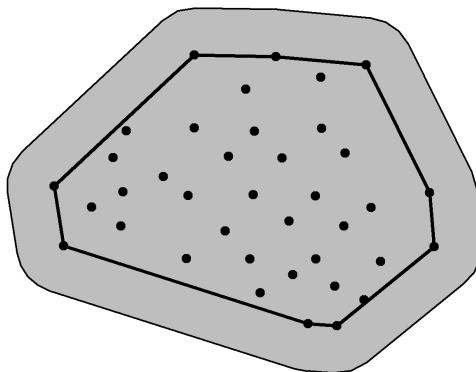


Figure 3.2: Convex hull of the bear hair snare array buffered by mean female home range radius (2.19 km).

```

413 plot(p1ch, border='black', lwd=2, add=TRUE)
414 gArea(bp1)
415 }
416
417 bcharea(2.19,traplocs=trapmat)

```

418 The resulting buffered convex hull is shown in Fig. 3.2.

419 To conjure up a density estimate under model  $M_0$ , we compute the appropriate posterior summary of  $N$  and the prescribed area ( $277.011 \text{ km}^2$ ):

```

421 > summary(fit0$sims.list$N/277.011)
422   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
423 0.1697 0.1733 0.1805 0.1803 0.1841 0.2130
424
425 > quantile(fit0$sims.list$N/277.011,c(0.025,0.975))
426    2.5%    97.5%
427 0.1696684 0.1949381

```

428 which yields a density estimate of about  $0.18 \text{ ind/km}^2$ , and a 95% Bayesian confidence interval of  $(0.170, 0.195)$ .

430 In summary, we have an estimate of density if we have faith in our stated value of the “sample area”. Clearly though this is largely subjective, and not something we can formally evaluate from the data. How certain are we of this area? Can we quantify our uncertainty about this quantity? More important,

what exactly is the meaning of this area and, in this context, how do we gauge bias and/or variance of “estimators” of it? (i.e., what is it estimating?).<sup>3</sup> There is no theory to guide us in trying to answer these important questions.

### 3.3 Temporally varying and behavioral effects

The purpose of this chapter is mainly to emphasize the central importance of the binomial model in capture-recapture and so we have considered models for individual encounter frequencies - the number of times individuals are captured out of  $K$  samples. Sometimes it is not acceptable to aggregate the encounter data for each individual - such as when encounter probability varies over time among samples. Time-varying responses that are relevant in many capture-recapture studies are “effort” such as amount of search time, number of observers, or trap nights, or when encounter probability varies over time or as a function of date or season due to species behavior (Kéry et al., 2010). A common situation in many animal studies is that in which there exists a “behavioral response” to trapping (even if the animal is not physically trapped).

Behavioral response is an important concept in animal studies because individuals might learn to come to baited traps or avoid traps due to trauma related to being encountered. There are a number of ways to parameterize a behavioral response to encounter. The distinction between persistent and ephemeral was made by Yang and Chao (2005) who considered a general behavioral response model of the form:

$$\text{logit}(p_{ik}) = \alpha_0 + \alpha_1 * y_{i,k-1} + \alpha_2 x_{ik}$$

where  $x_{ik}$  is a covariate indicator variable of previous capture (i.e.,  $x_{ik} = 1$  if captured in any previous period). Therefore, encounter probability changes depending on whether an individual was captured in the immediate previous period (ephemeral behavioral response XXX described by the term  $\alpha_1 * y_{i,k-1}$ ) or in any previous period (persistent behavioral response). The former probably models a behavioral response due to individuals moving around their territory relatively slowly over time and the latter probably accommodates trap happiness due to baiting or shyness due to trauma. Spatial capture-recapture models allow us to include trap-specific covariates, and in such models it makes sense to consider a local behavioral response that is trap-specific (Royle et al., 2011b) - that is, the encounter probability is modified for an individual trap depending on previous capture in that trap.

Models with temporal effects are easy to describe in the **BUGS** language and analyze and we provide a number of examples in Chapt. 9 and elsewhere.

---

<sup>3</sup>IS it worth MENTIONING THE DELTA APPROXIMATION KARANTH AND NICHOLS (1998)?

### 3.4 Models with individual heterogeneity

Here we consider models with individual-specific encounter probability parameters, say  $p_i$ , which we model according to some probability distribution,  $g(\theta)$ . We denote this basic model assumption as  $p_i \sim g(\theta)$ . This type of model is similar in concept to extending a GLM to a GLMM but in the capture-recapture context  $N$  is unknown. The basic class of models is often referred to as “model  $M_h$ ” but really this is a broad class of models, each being distinguished by the specific distribution assumed for  $p_i$ . There are many different varieties of model  $M_h$  including parametric and various putatively non-parametric approaches (Burnham and Overton, 1978; Norris III and Pollock, 1996; Pledger, 2000). One important practical matter is that estimates of  $N$  can be extremely sensitive to the choice of heterogeneity model (Fienberg et al., 1999; Dorazio and Royle, 2003; Link, 2003). Indeed, Link (2003) showed that in some cases it’s possible to find models that yield precisely the same expected data, yet produce wildly different estimates of  $N$ . In that sense,  $N$  for most practical purposes is not identifiable across classes of mixture models, and this should be understood before fitting any such model. One solution to this problem is to seek to model explicit factors that contribute to heterogeneity, e.g., using individual covariate models (See 3.5 below). Indeed, spatial capture-recapture models seek to do just that, by modeling heterogeneity due to the spatial organization of individuals in relation to traps or other encounter mechanism. For additional background and applications of model  $M_h$  see Royle and Dorazio (2008, Chapt. 6) and Kéry and Schaub (2011, Chapt. 6).

Model  $M_h$  has important historical relevance to spatial capture-recapture situations (Karanth, 1995) because investigators recognized that the juxtaposition of individuals with the array of trap locations should yield heterogeneity in encounter probability, and thus it became common to use some version of model  $M_h$  in spatial trapping arrays to estimate  $N$ . While this doesn’t resolve the problem of not knowing the area relevant to  $N$ , it does yield an estimator that accommodates the heterogeneity in  $p$  induced by the spatial aspect of capture-recapture studies.

To see how this juxtaposition induces heterogeneity, we have to understand the relevance of movement in capture-recapture models. Imagine a quadrat that can be uniformly searched by a crew of biologists for some species of reptile (see Royle and Young (2008)). Figure 3.3 shows a sample quadrat searched repeatedly over a period of time. Further, suppose that species exhibits some sense of spatial fidelity in the form of a home range or territory, and individuals move about their home range (home range centroids are given by the blue dots) in some kind of random fashion. Heuristically, we imagine that each individual in the vicinity of the study area is liable to experience variable exposure to encounter due to the overlap of its home range with the sampled area - essentially the long-run proportion of times the individual is within the sample plot boundaries, say  $\phi$ . We might model the exposure of an individual to capture by supposing that  $z_i = 1$  if individual  $i$  is available to be captured (i.e., within the survey plot) during any sample, and 0 otherwise. Then,  $\Pr(z_i = 1) = \phi$ . In the

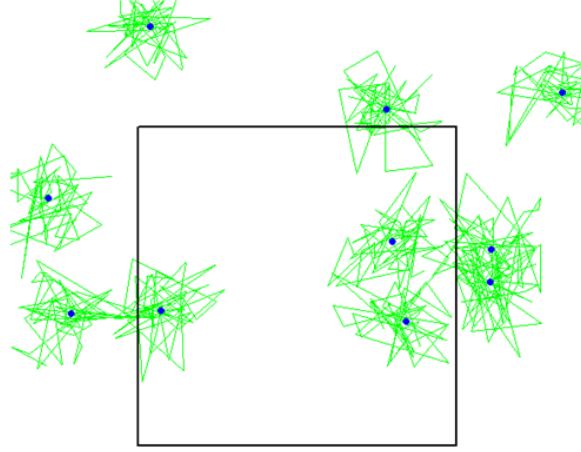


Figure 3.3: A quadrat searched for lizards and the locations of each lizard over some period of time.

context of spatial studies, it is natural that  $\phi$  should depend on *where* an individual lives, i.e., it should be individual-specific  $\phi_i$  (Chandler et al., 2011). This system describes, precisely, that of “random temporary emigration” (Kendall et al., 1997) where  $\phi_i$  is the individual-specific probability of being “available” for capture.

Conceptually, SCR models aim to deal with this problem of variable exposure to sampling due to movement in the proximity of the trapping array explicitly and formally with auxiliary spatial information. If individuals are detected with probability  $p_0$ , *conditional* on  $z_i = 1$ , then the marginal probability of detecting individual  $i$  is

$$p_i = p_0 \phi_i$$

so we see clearly that individual heterogeneity in encounter probability is induced as a result of the juxtaposition of individuals (i.e., their home ranges) with the sample apparatus and the movement of individuals about their home range.

We will work with a specific type of model  $M_h$  here, that in which we extend the basic binomial observation model of model  $M_0$  so that

$$\text{logit}(p_i) = \mu + \eta_i$$

where

$$\eta_i \sim \text{Normal}(0, \sigma_p^2)$$

531 We could as well write

$$\text{logit}(p_i) \sim \text{Normal}(\mu, \sigma_p^2)$$

532 This “logit-normal mixture” was analyzed by Coull and Agresti (1999) and  
 533 elsewhere. It is a natural extension of the basic model with constant  $p$ , as  
 534 a mixed GLMM, and similar models occur throughout statistics. It is also  
 535 natural to consider a beta prior distribution for  $p_i$  (Dorazio and Royle, 2003)  
 536 and so-called “finite-mixture” models (models in which individuals are assumed  
 537 to belong to a finite number of latent classes, each of which has its own capture  
 538 probability) are also popular (Norris III and Pollock, 1996; Pledger, 2000).

### 539 3.4.1 Analysis of Model $M_h$

540 If  $N$  is known, it is worth taking note of the essential simplicity of model  $M_h$  as  
 541 a binomial GLMM. This is a type of model that is widely applied in throughout  
 542 statistics using standard methods of inference based either on integrated likeli-  
 543 hood (Laird and Ware, 1982; Berger et al., 1999) which we discuss in Chapt. 6  
 544 or standard Bayesian methods. However, because  $N$  is not known, inference is  
 545 somewhat more challenging. We address that here using Bayesian analysis based  
 546 on data augmentation (DA). Although we use data augmentation in the context  
 547 of Bayesian methods here, we note that heterogeneity models formulated under  
 548 DA are easily analyzed by conventional likelihood methods as zero-inflated bi-  
 549 nomial mixtures (Royle, 2006) and more traditional analysis of model  $M_h$  based  
 550 on integrated likelihood, without using data augmentation, has been considered  
 551 by Coull and Agresti (1999), Dorazio and Royle (2003), and others.

552 As with model  $M_0$ , we have the Bernoulli model for the zero-inflation vari-  
 553 ables:  $z_i \sim \text{Bern}(\psi)$  and the model of the observations expressed conditional on  
 554 the latent variables  $z_i$ . For  $z_i = 1$ , we have a binomial model with individual-  
 555 specific  $p_i$ :

$$y_i | z_i = 1 \sim \text{Bin}(K, p_i)$$

556 and otherwise  $y_i | z_i = 0 \sim \delta(0)$ . Further, we prescribe a distribution for  $p_i$ . Here  
 557 we assume

$$\text{logit}(p_i) \sim \text{Normal}(\mu, \sigma^2)$$

558 The basic **BUGS** description for this model, assuming a  $\text{Unif}(0, 1)$  prior for  
 559  $p_0 = \text{logit}^{-1}(\mu)$ , is given as follows:

```
560 model{
561
562   p0 ~ dunif(0,1)          # prior distributions
563   mup<- log(p0/(1-p0))
564   taup~dgamma(.1,.1)
565   psi~dunif(0,1)
566
567   for(i in 1:(nind+nz)){
568     z[i]~dbern(psi)        # zero inflation variables
569     lp[i] ~ dnorm(mup,taup) # individual effect
```

```

570   logit(p[i])<-lp[i]
571   mu[i]<-z[i]*p[i]
572   y[i]~dbin(mu[i],J) # observation model
573 }
574
575 N<-sum(z[1:(nind+nz)]) # N is a derived parameter
576 }

```

### 577 3.4.2 Analysis of the Fort Drum data

578 The logit-normal heterogeneity model was fitted to the bear data from the Fort  
579 Drum study, and we used data augmentation to produce a data set of  $M = 500$   
580 individuals. We ran the model using **JAGS** with the instructions given as  
581 follows:

```

582 [... get data as before ....]
583
584 set.seed(2013)
585
586 cat("
587 model{
588   p0 ~ dunif(0,1)          # prior distributions
589   mup<- log(p0/(1-p0))
590   sigmap ~ dunif(0,10)
591   taup<- 1/(sigmap*sigmap)
592   psi~dunif(0,1)
593
594   for(i in 1:(nind+nz)){
595     z[i]~dbern(psi)        # zero inflation variables
596     lp[i] ~ dnorm(mup,taup) # individual effect
597     logit(p[i])<-lp[i]
598     mu[i]<-z[i]*p[i]
599     y[i]~dbin(mu[i],K)    # observation model
600   }
601
602   N<-sum(z[1:(nind+nz)])
603 }
604 ",file="modelMh.txt")
605
606 data1<-list(y=ytot, nz=nz, nind=nind,K=K)
607 params1= c('p0','sigmap','psi','N')
608 inits = function() {list(z=as.numeric(ytot>=1), psi=.6, p0=runif(1),
609   sigmap=runif(1,.7,1.2),lp=rnorm(M,-2)) }
610
611 library("rjags")
612 jm<- jags.model("modelMh.txt", data=data1, inits=inits, n.chains=4,
613   n.adapt=1000)
614 jout<- coda.samples(jm, params1, n.iter=200000, thin=1)

```

615 This produces the posterior distribution for  $N$  shown in Fig. 3.4. Posterior  
616 summaries of parameters are given as follows:

```

617 > summary(jout)
618
619 Iterations = 2001:202000
620 Thinning interval = 1
621 Number of chains = 4
622 Sample size per chain = 2e+05
623
624 1. Empirical mean and standard deviation for each variable,
625    plus standard error of the mean:
626
627           Mean      SD Naive SE Time-series SE
628 N      117.7740 56.31633 6.296e-02      1.960115
629 p0       0.0728  0.05522 6.174e-05      0.001655
630 psi      0.2366  0.11362 1.270e-04      0.003909
631 sigmap   2.0795  0.53096 5.936e-04      0.016789
632
633 2. Quantiles for each variable:
634
635           2.5%      25%      50%      75%      97.5%
636 N      62.000000 82.00000 102.00000 134.0000 277.0000
637 p0       0.003143  0.02842  0.06077  0.1066  0.2036
638 psi      0.117269  0.16377  0.20522  0.2712  0.5560
639 sigmap   1.211900  1.69434  2.02113  2.4028  3.2694

```

640 andy stopped here

641 We used  $M = 500$  for this analysis and we note that while the posterior  
642 mass of  $N$  is concentrated away from this upper bound (Fig. 3.4), the posterior  
643 has an extremely long right tail, with some posterior values at the upper bound  
644  $N = 500$ . Maybe or maybe not sufficient data augmentation.<sup>4</sup> The model runs  
645 effectively in **WinBUGS** but sometimes with apparently inefficient mixing for  
646 reasons that may be related to bad starting values. In some cases this was  
647 resolved if we supplied starting values for the  $\text{logit}(p_i)$  parameters and  $\tau$ .

648 XX  
649 ANDY NEED DENSITY ESTIMATES AFTER PUTTING RESULTS for 500k  
650 XX

651 Because of the skewed posterior we see that the posterior mean ( $N = 117$ )  
652 is considerably higher than the posterior mode ( $N = 102$ ). Moreover, posterior  
653 summaries are estimated with a relatively high error: The “Time-series” or  
654 Monte Carlo SE of around 2.0 (see secs. for discussion of this quantity ?? ??).  
655 Further, it may be surprising that the posterior mode does not compare well  
656 with the MLE. To compute the posterior mode we could easily find the posterior  
657 value of  $N$  with the highest mass because  $N$  is discrete. But we want to smooth  
658 out some of the Monte Carlo error a bit so we used a smoothing spline to the  
659 posterior frequencies of  $N$  as follows:

```

660 tt<-table(jout[[1]][, "N"])[1:80]
661 xg<-as.numeric(names(tt))

```

---

<sup>4</sup> **to do:** insert final results. longer run. more data augmentation. compare with winbugs.



```

662 plot(xg,tt)
663 sp<- smooth.spline(xg,tt,df=9)
664 sp$x[sp$y==max(sp$y)]
665 [1] 80

```

666 The `df` argument controls the degree of smoothing and we find in this case that  
 667 the modal value (i.e., 80) is not too sensitive to the smoothing parameter but  
 668 this should be checked in any specific instance<sup>5</sup>.

669 To compute the MLE, we used the **R** code contained in Panel 6.1 of Royle  
 670 and Dorazio (2008). The MLE of  $\log(n_0)$ , the logarithm of the number of  
 671 uncaptured individuals, is  $\widehat{\log(n_0)} = 3.86$  and therefore  $\hat{N} = \exp(3.86) + 47 =$   
 672  $94.47$  which is not at all consistent with the apparent mode in Fig. 3.4.<sup>6</sup>

673 **Remarks:** First of all the posterior for this model and data set is very sen-  
 674 sitive to prior distributions. While MLEs are invariant to transformation of the  
 675 parameters, the posterior distribution definitely is *not* invariant. In the present  
 676 case, the use of a  $\text{Unif}(0, 1)$  prior for  $p_0 = \text{expit}(\mu)$  is somewhat informative – in  
 677 particular, it is not at all “flat” on the scale of  $\mu$  – and this affects the posterior.  
 678 We generally always recommend use of a  $\text{Unif}(0, 1)$  prior for  $\text{expit}(\mu)$  in such  
 679 models. That said, we were surprised at this result, and we experimented with  
 680 other prior configurations including putting a flat prior on  $\mu$  directly. That  
 681 specific prior suggests the possibility that the posterior distribution may be  
 682 improper for that prior specification. This kind of small sample instability has  
 683 been widely noted in model  $M_h$  (Fienberg et al., 1999; Dorazio and Royle, 2003),  
 684 as has extreme sensitivity to the specific form of model  $M_h$  (Link, 2003). In  
 685 summary, while the mode is well-defined, the data set is relatively sparse and  
 686 hence inferences are poor and sensitive to model choice.

687 **Building your own MCMC algorithm:** For fun, we construct our own  
 688 MCMC algorithm using a Metropolized Gibbs sampler for model  $M_h$  in Chapt.  
 689 7, where we also develop the MCMC algorithms for spatial capture-recapture  
 690 models.

## 691 3.5 Individual Covariate Models: Toward Spa- 692 tial Capture-Recapture

693 A standard situation in capture-recapture models is when an individual covari-  
 694 ate is measured, and this covariate is thought to influence encounter probability.  
 695 As with other closed population models, we begin with the basic binomial ob-  
 696 servation model:

$$y_i \sim \text{Bin}(K, p_i)$$

<sup>5</sup>we need to give examples of using `density()` to obtain modes

<sup>6</sup>We note that the result is inconsistent with Gardner et al. (2009) who reported an MLE of 104.1 ( $\text{density} = 0.437 \text{ inds}/\text{km}^2$ ) although we do not know the reason for this at the present time.

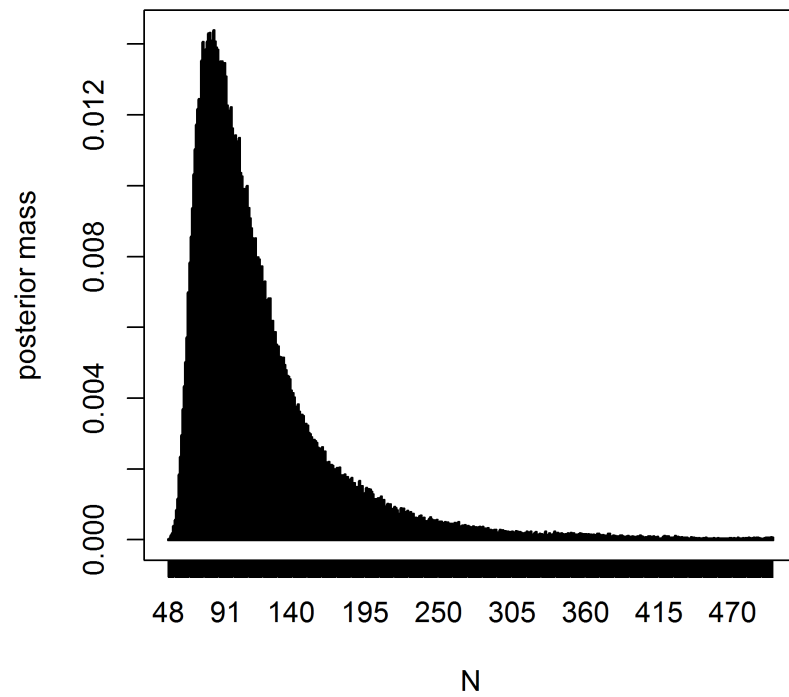


Figure 3.4: Posterior of  $N$  for Fort Drum bear study data under the logit-normal version of model  $M_h$ .

and we assume also a model for encounter probability according to:

$$\text{logit}(p_i) = \alpha + \beta x_i \quad (3.5)$$

Classical examples of covariates influencing detection probability are type of animal (juvenile/adult or male/female), a continuous covariate such as body mass (Royle and Dorazio, 2008, ch. 6), or a discrete covariate such as group or cluster size. For example, in models of aerial survey data, it is natural to model detection probabilities as a function of the observation-level individual covariate, “group size” (Royle, 2008, 2009; Langtimm et al., 2011).

Such “individual covariate models” are similar in structure to model  $M_h$ , except that the individual effects are *observed* for the  $n$  individuals that appear in the sample. These models are important here because spatial capture-recapture models are precisely a form of individual covariate model, an idea that we will develop here and elsewhere. Specifically, they are such models, but where the individual covariate is a partially observed latent variable for captured individuals. As such, it is a type of measurement error. That is, unlike model  $M_h$ , we do have some direct information about the latent variable, which comes from the spatial locations/distribution of individual recaptures.

Traditionally, estimation of  $N$  in individual covariate models is achieved using methods based on ideas of unequal probability sampling (i.e., Horwitz-Thompson estimation; see Huggins (1989) and Alho (1990)). An estimator of  $N$  is

$$\hat{N} = \sum_i^n \frac{1}{\tilde{p}_i}$$

where  $\tilde{p}_i$  is the probability that individual  $i$  appeared in the sample. That is,  $\tilde{p}_i = \Pr(y_i > 0)$  where, in closed population capture-recapture models,

$$\Pr(y_i > 0) = (1 - (1 - p_i)^K)$$

where  $p_i$  is a function of parameters  $\alpha$  and  $\beta$  according to Eq. 3.5. In practice, parameters are estimated from the conditional-likelihood of the observed encounter histories which is, for observation  $y_i$ ,

$$\mathcal{L}_c(\alpha, \beta | y_i) = \frac{\text{Bin}(y_i | \alpha, \beta)}{\tilde{p}_i}.$$

Here we take a formal model-based approach to Bayesian analysis of such models based on the joint likelihood using data augmentation (Royle, 2009). Classical likelihood analysis of the so-called “full likelihood” is covered by Borchers et al. (2002). For Bayesian analysis of individual covariate models, because the individual covariate is unobserved for the  $N - n$  uncaptured individuals, we require a model to describe variation among individuals, essentially allowing the sample to be extrapolated to the population. For our present purposes, we consider a continuous covariate and we assume that it has a normal distribution:

$$x_i \sim \text{Normal}(\mu, \sigma^2)$$

Data augmentation can be applied directly to this class of models. In particular, reformulation of the model under DA yields a basic zero-inflated binomial model of the form:

$$\begin{aligned} z_i &\sim \text{Bern}(\psi) \quad i = 1, 2, \dots, M \\ y_i | z_i = 1 &\sim \text{Bin}(K, p_i(x_i)) \\ y_i | z_i = 0 &\sim \delta(0) \\ x_i &\sim \text{Normal}(\mu, \sigma^2) \end{aligned}$$

Fully spatial capture-recapture models use this formulation with a latent covariate that is directly related to the individual detection probability (see next section). As with the previous models, implementation is trivial in the **BUGS** language. The **BUGS** specification is very similar to that for model  $M_h$ , but we require the distribution of the covariate to be specified, along with priors for the parameters of that distribution.

### 3.5.1 Example: Location of capture as a covariate.

If we had a regular grid of traps over some closed geographic system then we imagine that the average location of capture would be a decent estimate (heuristically) of an individual's home range center. Intuitively some measure of typical distance from home range center to traps for an individual should be a decent covariate to explain heterogeneity in encounter probability, i.e., individuals with more exposure to traps should have higher encounter probabilities and vice versa. A version of this idea was put forth by Boulanger and McLellan (2001) (see also Ivan (2012)), but using the Huggins-Alho estimator and with covariate "distance to edge" of the trapping array. A limitation of this approach is that it does not provide a solution to the problem that the trap area is fundamentally ill-defined, nor does it readily accommodate the inherent and heterogeneous variation in this measured covariate.

Here, we provide an example of this type of heuristically motivated approach using the fully model-based individual covariate model described above analyzed by data augmentation. We take a slightly different approach than that adopted by Boulanger and McLellan (2001). By analyzing the full likelihood and placing a prior distribution on the individual covariate, we resolve the problem of having an ill-defined area over which the population size is distributed. After you read later chapters of this book, it will be apparent that SCR models represent a formalization of this heuristic procedure.

For our purposes here, we define  $x_i = \|\mathbf{s}_i - \mathbf{x}_0\|$  where  $\mathbf{s}_i$  is the average encounter location of individual  $i$  and  $\mathbf{x}_0$  is the centroid of the trap array. Conceptually, individuals in the middle of the array should have higher probability of encounter and, as  $x_i$  increases,  $p_i$  should therefore decrease. We note that we have defined  $\mathbf{s}_i$  in terms of a sample quantity - the observed mean - which is ad hoc but consistent with existing applications in the literature. For an expansive, dense trapping grid then we might expect the sample mean encounter location to be a good estimate of home range center but, clearly this is biased

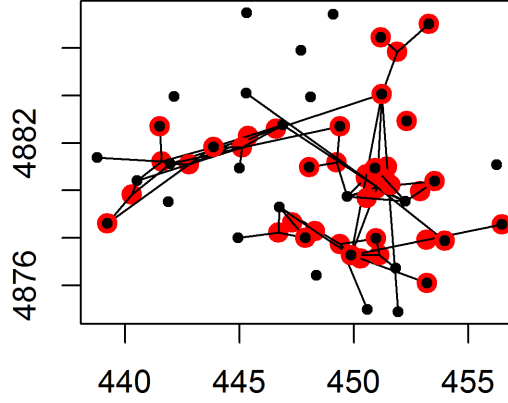


Figure 3.5: Spider plot of the Fort Drum study data.

768 for individuals that live around the edge (or off) the trapping array. Regardless,  
 769 it should be good enough for our present purposes of demonstrating this heuris-  
 770 tically appealing application of an individual covariate model. A key point is  
 771 that  $\mathbf{s}_i$  is missing for each individual that is not encountered and thus so is  $x_i$ .  
 772 Thus, it is a latent variable, or random effect, and we need therefore to specify  
 773 a probability distribution for it. As a measurement of distance we know it must  
 774 be positive-valued. Thinking about this like a distance sampling problem lets  
 775 first try to make  $x_i$  uniform from 0 to some large number, say  $D_{max}$ , beyond  
 776 which it would be difficult to imagine an individual being captured. For exam-  
 777 ple,  $D_{max}$  should be at a home range diameter past the furthest trap from the  
 778 center. As such, we use this distribution for the individual covariate “distance  
 779 from center of the trap array”

$$x_i \sim \text{Unif}(0, D_{max})$$

780 where  $D_{max}$  is a specified constant, which we may choose to be arbitrarily large.  
 781 In practice, people have used distance from edge of the trap array but that is  
 782 less easy to make sense of.

783 **Fort Drum Bear Study**

784 We have to do a little bit of data processing to fit this individual covariate  
 785 model to the Fort Drum data. We need to compute the individual covariate  
 786  $\mathbf{x}_i$  (distance from the centroid of the trapping array) using the **R** function  
 787 **spiderplot** provided in **scrbook**. This function also produces the keen plot  
 788 shown in Fig. 3.5 which we call a “spider plot”. The **R** commands for obtaining  
 789 the individual covariate “distance from trap centroid” and making the spider  
 790 plot are as follows:

```
791 library("scrbook")
792 data("beardata")
793 toad<- spiderplot(beardata$bearArray,beardata$trapmat)
794 xcent<-toad$xcent
```

795 For the analysis of these data using the individual covariate “distance from  
 796 centroid” we used  $x_i \sim \text{Unif}(0, D_{max})$  with  $D_{max} = 11.5 \text{ km}^2$  which is about  
 797 the distance from the array center to the furthest trap. Once we pick  $D_{max}$   
 798 then the direct implication is that the population size parameter applies to the  
 799 area within 11.5 units of the trap centroid and thus we will find that  $N$  does,  
 800 in fact, scale with our choice of  $D_{max}$  to reflect the changing area over which  
 801 the  $N$  individuals of the model reside. The **BUGS** model specification and **R**  
 802 commands to package the data and fit the model are as follows:

```
803 cat("
804 model{
805   p0 ~ dunif(0,1)          # prior distributions
806   mup<- log(p0/(1-p0))
807   psi~dunif(0,1)
808   beta~dnorm(0,.01)
809
810   for(i in 1:(nind+nz)){
811     xcent[i]~dunif(0,Dmax)
812     z[i]~dbern(psi)        # DA variables
813     lp[i] <- mup + beta*xcent[i] # individual effect
814     logit(p[i])<-lp[i]
815     mu[i]<-z[i]*p[i]
816     y[i]~dbin(mu[i],K)    # observation model
817   }
818   N<-sum(z[1:(nind+nz)])
819 }
820 ",file="modelMcov.txt")
821
822 data2<-list(y=ytot,nz=nz,nind=nind,K=K,xcent=xcent,Dmax=11.5)
823 params2<-list('p0','psi','N','beta')
824 inits = function() {list(z=zst, psi=psi, p0=runif(1),beta=rnorm(1) ) }
825 fit2 = bugs(data2, inits, params2, model.file="modelMcov.txt",working.directory=getwd(),
826             debug=T, n.chains=3, n.iter=11000, n.burnin=1000, n.thin=1)
```

827 This produces the following posterior summary statistics:

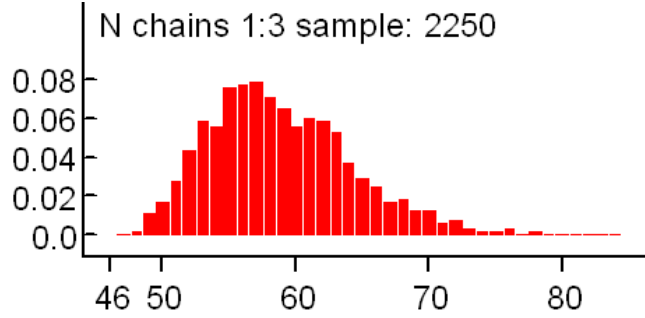


Figure 3.6: Needs a caption

```

828 Inference for Bugs model at "modelMcov.txt", fit using WinBUGS,
829 3 chains, each with 11000 iterations (first 1000 discarded)
830 n.sims = 30000 iterations saved
831      mean      sd    2.5%    25%    50%    75%   97.5%  Rhat  n.eff
832 p0         0.54  0.07    0.40    0.50    0.54    0.59    0.67    1   1100
833 psi        0.34  0.05    0.25    0.31    0.34    0.37    0.44    1   3500
834 N         58.92  5.49   50.00   55.00   58.00   62.00   71.00    1   1900
835 beta      -0.25  0.06   -0.36   -0.29   -0.25   -0.21   -0.12    1    780
836 deviance 459.51 13.21 435.80 450.20 458.80 467.90 487.40    1  2600

```

837 It might be perplexing that the estimated  $N$  is much lower than obtained  
838 by model  $M_h$  but there is a good explanation for this, discussed subsequently.  
839 That issue notwithstanding, it is worth pondering how this model could be  
840 an improvement (conceptually or technically) over some other model/estimator  
841 including  $M_0$  and  $M_h$  considered previously. Well, for one, we have accounted  
842 formally for heterogeneity due to spatial location of individuals relative to expo-  
843 sure to the trap array, characterized by the centroid of the array. Moreover, we  
844 have done so using a model that is based on an explicit mechanism, as opposed  
845 to a phenomenological one such as Model  $M_h$ . Moreover, importantly, using our  
846 new model, *the estimated  $N$  applies to an explicit area which is defined by our*  
847 *prescribed value of  $D_{max}$ .* That is, this area is a fixed component of the model  
848 and the parameter  $N$  therefore has explicit spatial context, as the number of  
849 individuals with home range centers less than  $D_{max}$  from the centroid of the  
850 trap array. As such, the implied “effective area” of the trap array for a given  
851  $D_{max}$  is a precisely defined quantity – it is that of a circle with with radius  
852  $D_{max}$ .

### 853 3.5.2 Extension of the Model

854 This model is actually not a very good model for one important reason: Impos-  
855 ing a uniform prior distribution on  $x$  implies that density is *not constant* over  
856 space. In particular, this model implies that it *decreases* as we move away from

the centroid of the trap array. That is,  $x_i \sim \text{Unif}(0, D_{max})$  implies constant  $N$  in each distance band from the centroid but obviously the *area* of each distance band is increasing. This is one reason we have a lower estimate of density than that obtained previously from model  $M_0$  and also why, if we were to increase  $D_{max}$ , we would see density continue to decrease.

Fortunately, the use of an individual covariate model is *not* restricted to use of this specific distribution for the individual covariate. Clearly, it is a bad choice and, therefore, we should think about whether we can choose a better distribution for  $D_{max}$  - one that doesn't imply a decreasing density as distance from the centroid increases. Conceptually, what we want to do is impose a prior on distance from the centroid,  $x$ , such that density is proportional to the amount of area in each successive distance band as you move farther away from the centroid. In fact, there is theory that exists which tells us what the correct distribution of  $x$  is  $2x/D_{max}^2$ . This can be derived by noting that  $F(x) = \Pr(X < x) = \pi * x * x / \pi * D_{max}^2$ . Then,  $f(x) = dF/dx = 2 * x / (D_{max}^2)$ . This is a sort of triangular distribution in density induced because the incremental area in each additional distance band increases linearly with radius (i.e., distance from centroid). It is sometimes comforting to verify things empirically:

```

875 u<-runif(10000,-1,1)
876 v<-runif(10000,-1,1)
877 d<- sqrt(u*u+v*v)
878 hist(d[d<1])
879 hist(d[d<1],100)
880 hist(d[d<1],100,probability=TRUE)
881 abline(0,2)

```

It would be useful if we could describe this distribution in **BUGS** but there is not a built-in way to do this that we are aware of. One possibility is to use a discrete version of the pdf. We might also be able to use what is referred to in **WinBUGS** jargon as the “zeros trick” (see *Advanced BUGS tricks* in the manual) although we haven't pursued this approach. Instead, we use a discrete approximation of the density of  $x$ , and break  $D_{max}$  into  $L$  distance classes of width  $\delta$ , with probabilities proportional to  $2 * x$ . In particular, if we denote the cut-points by  $xg_1 = 0, xg_2, \dots, xg_{L+1} = D_{max}$  and the interval midpoints are  $xm_i = xg_{i+1} - \delta$  then the interval probabilities are  $p_i = 2 * xm_i * \delta / (D_{max}^2)$ , which we can compute once and then pass them to **WinBUGS** as data.

The **R** commands for doing all of this (noting that we have already loaded and processed the Fort Drum bear data) are given as follows. In the model description the variable  $x$  (observed distance from centroid of the trap array) has been rounded so that the discrete version of the  $f(x)$  can be used as described previously. The new variable labeled **xround** is then the integer category label in units of  $\delta$  from 0. Thus, to convert back to distance in the expression for  $lp[i]$ , **xround[i]** has to be multiplied by  $\delta$ . Here is the **BUGS** model specification:

```

899 delta<-0.2
900 xround<-xcent/%delta + 1
901 Dgrid<- seq(delta,Dmax,delta)
902 xprobs<- delta*(2*Dgrid/(Dmax*Dmax))
903 xprobs<-xprobs/sum(xprobs)

```



```

904
905 cat("
906 model{
907   p0 ~ dunif(0,1)          # prior distributions
908   mup<- log(p0/(1-p0))
909   psi~dunif(0,1)
910   beta~dnorm(0,.01)
911
912   for(i in 1:(nind+nz)){
913     xround[i]~dcat(xprobs[])
914     z[i]~dbern(psi)         # zero inflation variables
915     lp[i] <- mup + beta*xround[i]*delta # individual effect
916     logit(p[i])<-lp[i]
917     mu[i]<-z[i]*p[i]
918     y[i]~dbin(mu[i],K)     # observation model
919   }
920
921   N<-sum(z[1:(nind+nz)])
922 }
923 ",file="modelMcov.txt")

```

924 To fit the model we do this - keeping in mind that the data objects required
925 below have been defined in previous analyses of this chapter:

```

926 data2<-list(y=ytot,nz=nz,nind=nind,K=K,xround=xround,xprobs=xprobs,delta=delta)
927 params2<-list('p0','psi','N','beta')
928 inits = function() {list(z=z, psi=psi, p0=runif(1),beta=rnorm(1) ) }
929 fit = bugs(data2, inits, params2, model.file="modelMcov.txt",
930           working.directory=getwd(), debug=FALSE, n.chains=3, n.iter=11000,
931           n.burnin=1000, n.thin=2)

```

932 This is a useful model because it induces a clear definition of area in which
933 the population of  $N$  individuals reside. Under this model, that area is defined
934 by specification of  $D_{max}$ . Further, the parameter  $N$  of the model is, explicitly,
935 the population size that applies to the particular value of  $D_{max}$  and, as such, we
936 will see that  $N$  scales with our choice of  $D_{max}$ . This might be disconcerting to
937 some – we can get whatever value of  $N$  we want by changing  $D_{max}$ ! Fortunately,
938 we find empirically, that while  $N$  seems highly sensitive to the prescribed value
939 of  $D_{max}$ , density seems to be invariant to  $D_{max}$  as long as it is chosen to
940 be sufficiently large. We fit the model for a random of values of  $D_{max}$  from
941  $D_{max} = 12$  (restricting values of  $x$  to be in close proximity to the trap array)
942 on up to 20. The results are given in Table 3.4.

943 andy stopped here

944 We see that the posterior mean and SD of density (individuals per square
945 km) appear insensitive to choice of  $D_{max}$  once we get a ways away from the
946 maximum observed value of about 11.5. The estimated density of 0.25 per
947 km<sup>2</sup> is actually quite a bit lower than we reported using model  $M_h$  for which
948 no relevant “area” quantity is explicit in the model. Using MLEs of  $N$  in
949 conjunction with buffer strips (see Table ??) our estimates were in the range of

Table 3.4: Analysis of Fort Drum bear hair snare data using the individual covariate model, for different values of  $D_{max}$ , the upper limit of the uniform distribution of ‘distance from centroid of the trap array’. “Density” is the posterior mean of density and SD is the posterior standard deviation.

| $D_{max}$ | Density | SD    |
|-----------|---------|-------|
| 12        | 0.230   | 0.038 |
| 15        | 0.244   | 0.041 |
| 17        | 0.249   | 0.044 |
| 18        | 0.249   | 0.043 |
| 19        | 0.250   | 0.043 |
| 20        | 0.250   | 0.044 |

0.32–0.43 and the Bayesian estimates were XXXX (posterior mode of  $N = 102$ ) or XXX (posterior mean of  $N = 117$ ) (see sec. 3.4 above). On the other hand our estimate of  $\hat{D} = 0.25$  here (based on the posterior mean) is higher than that reported from model  $M_0$  using the buffered area (0.18). There is no basis really for comparing or contrasting these various estimates and it would be a useful philosophical exercise for the reader to discuss this matter. In particular, application of models  $M_0$  and  $M_h$  are distinctly *not* spatially explicit models – the area within which the population<sup>7</sup> resides is not defined under either model. There is therefore no reason at all to think that the estimates produced under either either closed population model, based on a buffered “trap area”, are justifiable by any theory. In fact, we would get exactly the same estimate of  $N$  no matter what we declare the area to be. On the other hand, the individual covariate model explicitly describes a distribution for “distance from centroid” that is a reasonable and standard null model - it posits, in the absence of direct information, that individual home range centers are randomly distributed in space and that probability of detection depends on the distance between home range center and the centroid of the trap array. Under this definition of the system, we see that density is invariant to the choice of sample area which seems like a desirable feature.

The individual covariate model is not ideal, however, because it does not make full use of the spatial information in the data set, i.e., the trap locations and the locations of each individual encounter, and there is hope to extend this model in order to resolve remaining deficiencies.

### 3.5.3 Invariance of density to $D_{max}$

Under the model above, and also under models that we consider in later chapters, a general property of the estimators is that while  $N$  increases with the prescribed trap area (equivalent to  $D_{max}$  in this case), we expect that density estimators should be invariant to this area. In the model used above, we note

<sup>7</sup>We need to look back at Chapter 1 and make sure we quit calling this “sample area” - it really isn’t that at all, but rather the area within which  $N$  resides.

that  $Area(D_{max}) = \pi * D_{max}^2$  and  $E[N(D_{max})] = \lambda * Area(D_{max})$  and thus  $E[Density(D_{max})] = \lambda$ , i.e., constant. This should be interpreted as the *prior* density. Absent data, then realizations under the model will have density  $\lambda$  regardless of what  $D_{max}$  is prescribed to be. As we verified empirically above, the posterior density is also invariant Of  $D_{max}$  as long as the implied area is large enough so that the data no longer provide information about density (i.e., “far away”).

#### 3.5.4 Toward Fully Spatial Capture-recapture Models

We developed this model for the average observed location and equated it to home range center  $\mathbf{s}_i$ . Intuitively, taking the average encounter location as an estimate of home range center makes sense but more so when the trapping grid is dense and expansive relative to typical home range sizes. However, our approach also ignored the variable precision with which each  $\mathbf{s}_i$  is estimated and also, as noted previously, estimates of  $\mathbf{s}_i$  around the “edge” (however we define that) are biased because the observations are truncated (we can only observe locations within the trap array). In the next chapter we provide a further extension of this individual covariate model that definitively resolves the ad hoc nature of the individual covariate approach we took here. In that chapter we build a model in which  $\mathbf{s}_i$  are regarded as latent variables and the observation locations (i.e., trap specific encounters) are linked to those latent variables with an explicit model. We note that the model fitted previously could be adapted easily to deal with  $\mathbf{s}_i$  as a latent variable, simply by adding a prior distribution for  $\mathbf{s}_i$ . The reader should contemplate how to do this in **BUGS**.

### 3.6 DISTANCE SAMPLING: A primitive Spatial Capture-Recapture Model

Distance sampling is one of the most popular methods for estimating animal abundance. One of the great benefits of distance sampling is that it provides explicit estimates of *density*. The distance sampling model is a special case of a closed population model with a covariate. The covariate in this case,  $x_i$ , is the distance between an individual’s location “ $u$ ” and the observation location or transect. In fact, the model underlying distance sampling is precisely the same model as that which applies to the individual-covariate models, except that observations are made at only  $K = 1$  sampling occasion. In a sense, distance sampling is a spatial capture-recapture model, but without the “recapture.” This first and most basic spatial capture-recapture model has been used routinely for decades and, formally, it is a spatially-explicit model in the sense that it describes, explicitly, the spatial organization of individual locations (although this is not always stated explicitly) and, as a result, somewhat general models of how individuals are distributed in space can be specified (Royle et al., 2004; Johnson, 2010; Sillett, 2011).

As before, the distance sampling model, under data augmentation, includes a set of  $M$  zero-inflation variables  $z_i$  and the binomial model expressed conditional on  $z$  (binomial for  $z = 1$ , and fixed zeros for  $z = 0$ ). In distance sampling we pay for having only a single sample (i.e.,  $K = 1$ ) by requiring constraints on the model of detection probability. A standard model is

$$\log(p_i) = \beta x_i^2$$

for  $\beta < 0$ , where  $x_i$  denotes the distance at which the  $i$ th individual is detected relative to some reference location where perfect detectability ( $p = 1$ ) is assumed. This function corresponds to the “half-normal” detection function (i.e., with  $\beta = 1/\sigma^2$ ). If  $K > 1$  then an intercept in this model is identifiable and such models are usually called “capture-recapture distance sampling” (Alpizar-Jara and Pollock, 1996; Borchers et al., 1998).

As with previous examples, we require a distribution for the individual covariate  $x_i$ . The customary choice is

$$x_i \sim \text{Unif}(0, B)$$

wherein  $B > 0$  is a known constant, being the upper limit of data recording by the observer (i.e., the point count radius, or transect half-width). In practice, this is sometimes asserted to be infinity, but in such cases the distance data are usually truncated. Specification of this distance sampling model in the **BUGS** language is shown in Panel 3.2 from Royle and Dorazio (2008).

---

```

beta~dunif(0,10)
psi~dunif(0,1)

for(i in 1:(nind+nz)){
  z[i]~dbern(psi)      # DA Variables
  x[i]~dunif(0,B)      # B=strip width
  p[i]<-exp(logp[i])    # DETECTION MODEL
  logp[i]<-      - beta*(x[i]*x[i])
  mu[i]<-z[i]*p[i]
  y[i]~dbern(mu[i])    # OBSERVATION MODEL
}
N<-sum(z[1:(nind+nz)])
D<- N/striparea # area of transects

```

---

Panel 3.2: Distance sampling model in **BUGS**, using a half-normal detection function.

As with the individual covariate model in the previous section, the distance sampling model can be equivalently specified by putting a prior distribution on individual *location* instead of distance between individual and observation point

1039 (or transect). Thus we can write the general distance sampling model as

$$p_i = f(\beta, \|\mathbf{u}_i - \mathbf{x}_0\|)$$

1040 along with

$$\mathbf{u}_i \sim \text{Unif}(\mathcal{S})$$

1041 where  $\mathbf{x}_0$  is a fixed point (or line) and  $\mathbf{u}_i$  is the individual's location which is  
 1042 observable for  $n$  individuals. In practice it is easier to record distance instead  
 1043 of location. Basic math can be used to argue that if individuals have a uniform  
 1044 distribution in space, then the distribution of Euclidean distance is also uniform.  
 1045 In particular, if a transect of length  $L$  is used and  $x$  is distance to the transect  
 1046 then  $F(x) = \Pr(X \leq x) = L * x / L * B = x/B$  and  $f(x) = dF/dx = (1/B)$ . For  
 1047 measurements of radial distance, see the previous section.

1048 In the context of our general characterization of SCR models (Chapt. ??),  
 1049 we suggested that every SCR model can be described, conceptually, by a hier-  
 1050 archical model of the form:

$$[y|u][u|s][s].$$

1051 Distance sampling ignores the part of the model pertaining to  $\mathbf{s}$ , and deals only  
 1052 with the model components for the observed data  $\mathbf{u}$ <sup>8</sup>. Thus, we are left with a  
 1053 hierarchical model of the form

$$[y|\mathbf{u}][\mathbf{u}].$$

1054 In contrast, as we will see in the next chapters, basic SCR models (Chapt. 4)  
 1055 ignore  $\mathbf{u}$  and condition on  $\mathbf{s}$ , which is not observed:

$$[y|\mathbf{s}][\mathbf{s}]$$

1056 Since  $[\mathbf{u}]$  and  $[\mathbf{s}]$  are both assumed to be uniformly distributed, these are struc-  
 1057 turally equivalent models! The main differences have to do with interpretation  
 1058 of model components and whether or not the latent variables are observable (in  
 1059 distance sampling they are).

1060 So why bother with SCR models when distance sampling yields density  
 1061 estimates and accounts for spatial heterogeneity in detection? For one, imagine  
 1062 trying to collect distance sampling data on tigers! Clearly, distance sampling  
 1063 requires that one can collect large quantities of distance data, which is not always  
 1064 possible. For tigers, it is much easier, efficient, and safer to employ camera traps  
 1065 or tracking plates and then apply SCR models. Furthermore, as we will see in  
 1066 Chapt. ?? and ??, SCR models can use distance data to estimate all the  
 1067 parameters of our enchilada, allowing us to study distribution, movement, and  
 1068 density. Thus, SCR models are much more general and versatile than distance  
 1069 sampling models (which clearly are a special case), and can accommodate data  
 1070 from virtually all animal survey designs.

---

<sup>8</sup>Equivalently, we could also say that  $[u]$  in the distance sampling model is  $[u] = \int [u|\mathbf{s}][\mathbf{s}]d\mathbf{s}$

### 1071 3.6.1 Example: Muntjac deer survey from Nagarahole, In- 1072 dia

1073 Here we fit distance sampling models to distance sampling data on the muntjac  
1074 deer (*Muntiacus muntjak*) collected in the year 2004 from Nagarahole National  
1075 Park in southern India (Kumar et al. unpublished data). The muntjac is a  
1076 solitary species and distance measurements were made on 57 groups that were  
1077 largely singletons with 4 pairs of individuals. Commands for reading in and  
1078 organizing the data for **WinBUGS**, followed by writing the model to a text  
1079 file, are given below. Note that the total sampled area of the transects is fed  
1080 in as “striparea” which is 708 (km of transect walked) multiplied by the strip  
1081 width ( $B = 120 = 0.12$  km) multiplied by 2.

```
1082 library("R2WinBUGS")
1083 data<- read.csv("Muntjac.csv")
1084 hist(data[,3],30)
1085 nind<-nrow(data)
1086 y<-rep(1,nind)
1087 nz<-400
1088 y<-c(y,rep(0,nz))
1089 x<-data[,3]
1090 x<-c(x,rep(NA,nz))
1091 z<-y
1092
1093 cat("
1094 model{
1095   beta~dunif(0,10)
1096   psi~dunif(0,1)
1097
1098   for(i in 1:(nind+nz)){
1099     z[i]~dbern(psi)    # DA Variables
1100     x[i]~dunif(0,B)   # B=strip width
1101     p[i]<-exp(logp[i]) # DETECTION MODEL
1102     logp[i]<- -beta*(x[i]*x[i])
1103     mu[i]<-z[i]*p[i]
1104     y[i]~dbern(mu[i]) # OBSERVATION MODEL
1105   }
1106   N<-sum(z[1:(nind+nz)])
1107   D<- N/striparea # area of transects
1108 }
1109 ",file="dsamp.txt")
```

1110 Next, we provide inits, indicate which parameters to monitor, and then pass  
1111 those things to **WinBUGS**:

```
1112 data<-list(y=y,x=x,nz=nz,nind=nind,B=120,striparea=(708*2*.120))
1113 params<-list('beta','N','D','psi')
1114 inits = function() {list(z=z, psi=runif(1), beta=runif(1,0,.02) )}
1115 fit = bugs(data, inits, params, model.file="dsamp.txt",working.directory=getwd(),
1116           debug=T, n.chains=3, n.iter=11000, n.burnin=1000, n.thin=2)
```

Posterior summaries are provided in the following table. Estimated density is pretty low, 1.1 individuals per sq. km.<sup>9</sup>

```

Inference for Bugs model at "dsamp.txt", fit using WinBUGS,
3 chains, each with 11000 iterations (first 1000 discarded), n.thin = 2
n.sims = 15000 iterations saved

```

|          | mean   | sd    | 2.5%   | 25%    | 50%    | 75%    | 97.5%  | Rhat | n.eff |
|----------|--------|-------|--------|--------|--------|--------|--------|------|-------|
| beta     | 0.00   | 0.00  | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 1    | 1100  |
| N        | 185.73 | 26.53 | 138.00 | 167.00 | 184.00 | 203.00 | 242.00 | 1    | 570   |
| D        | 1.09   | 0.16  | 0.81   | 0.98   | 1.08   | 1.20   | 1.42   | 1    | 570   |
| psi      | 0.41   | 0.06  | 0.30   | 0.36   | 0.40   | 0.45   | 0.54   | 1    | 670   |
| deviance | 655.74 | 16.26 | 626.00 | 644.50 | 655.10 | 666.40 | 689.80 | 1    | 1300  |

```

[.... some output deleted .... ]

```

### 3.7 Summary and Outlook

Traditional closed population capture-recapture models are closely related to binomial generalized linear models. Indeed, the only real distinction is that in capture-recapture models, the population size parameter  $N$  (corresponding also to the size of a hypothetical “complete” data set) is unknown. This requires special consideration in the analysis of capture-recapture models. The classical approach to inference recognizes that the observations don’t have a standard binomial distribution but, rather, a truncated binomial (from which which the so-called “conditional likelihood” derives) since we only have encounter frequency data on observed individuals. If instead we analyze the models using data augmentation, the observations can be modeled using a zero-inflated binomial distribution. In short, when we deal with the unknown- $N$  problem using data augmentation then we are left with zero-inflated GLM and GLMMs instead of ordinary GLM or GLMMs. The analysis of such zero-inflated models is practically convenient, especially using the various Bayesian analysis packages that use the **BUGS** language.

Spatial capture-recapture models that we will consider in the rest of the chapters of this book are closely related to what have been called individual covariate models. Heuristically, spatial capture-recapture models arise by defining individual covariates based on observed locations of individuals – we can think of using some function of mean encounter location as an individual covariate. We did this in a novel way, by using distance to the centroid of the trapping array as a covariate. We analyzed the “full likelihood” using data augmentation, and placed a prior distribution on the individual covariate which was derived from an assumption that individual locations are, a priori, uniformly distributed in space. This assumption provides for invariance of the density estimator to the choice of population size area (induced by maximum distance from the centroid of the trap array). The model addressed some important problems in the use of

<sup>9</sup> This is much lower than Samba’s estimate produced from WinBUGS accounting for group size. Reason unknown.

1158 closed population models: it allows for heterogeneity in encounter probability  
1159 due to the spatial context of the problem and it also provides a direct estimate  
1160 of density because area is a feature of the model (via the prior on the individual  
1161 covariate). The model is still not completely general because it does not make  
1162 use of the fully spatial encounter histories, which provide direct information  
1163 about the locations and density of individuals. A specific individual covariate  
1164 model that is in widespread use is classical “distance sampling.” The model  
1165 underlying distance sampling is precisely a special kind of SCR model - but  
1166 one without replicate samples. Understanding distance sampling and individ-  
1167 ual covariate models more broadly provides a solid basis for understanding and  
1168 analyzing spatial capture-recapture models.



1169 **Chapter 4**

1170 **Fully Spatial**  
1171 **Capture-Recapture Models**



1172 **Chapter 5**

1173 **Other observation models**



1174 **Chapter 6**

1175 **Maximum likelihood**  
1176 **estimation**



1177 **Chapter 7**

1178 **MCMC details**





1179 **Chapter 8**

1180 **Goodness of Fit and stuff**



## 1181 Chapter 9

## 1182 Covariate models



## 1183 Chapter 10

# 1184 Inhomogeneous Point 1185 Process



## 1186 Chapter 11

## 1187 Open models





# Bibliography

- Alho, J. M. (1990), “Logistic regression in capture-recapture models,” *Biometrics*, 623–635.
- Alpizar-Jara, R. and Pollock, K. H. (1996), “A combination line transect and capture-recapture sampling model for multiple observers in aerial surveys,” *Environmental and Ecological Statistics*, 3, 311–327.
- Bales, S. L., Hellgren, E. C., Leslie Jr., D. M., and Hemphill Jr., J. (2005), “Dynamics of recolonizing populations of black bears in the Ouachita Mountains of Oklahoma,” *Wildlife Society Bulletin*, 1342–1351.
- Berger, J. O., Liseo, B., and Wolpert, R. L. (1999), “Integrated likelihood methods for eliminating nuisance parameters,” *Statistical Science*, 1–22.
- Borchers, D. L., Buckland, S. T., and Zucchini, W. (2002), *Estimating animal abundance: closed populations*, vol. 13, Springer Verlag.
- Borchers, D. L., Zucchini, W., and Fewster, R. M. (1998), “Mark-recapture models for line transect surveys,” *Biometrics*, 1207–1220.
- Boulanger, J. and McLellan, B. (2001), “Closure violation in DNA-based mark-recapture estimation of grizzly bear populations,” *Canadian Journal of Zoology*, 79, 642–651.
- Burnham, K. P. and Overton, W. S. (1978), “Estimation of the size of a closed population when capture probabilities vary among animals,” *Biometrika*, 65, 625.
- Chandler, R., Royle, J., and King, D. (2011), “Inference about density and temporary emigration in unmarked populations,” *Ecology*, 92, 1429–1435.
- Converse, S. J. and Royle, J. A. (2010), “Missing,” *Missing*, Missing.
- Coull, B. A. and Agresti, A. (1999), “The Use of Mixed Logit Models to Reflect Heterogeneity in Capture-Recapture Studies,” *Biometrics*, 55, 294–301.
- Dice, L. R. (1938), “Some census methods for mammals,” *Journal of Wildlife Management*, 2, 119–130.

- 1216 Dorazio, R. M. and Royle, J. A. (2003), "Mixture models for estimating the size  
1217 of a closed population when capture rates vary among individuals," *Biomet-*  
1218 *rics*, 351–364.
- 1219 Durban, J. and Elston, D. (2005), "Mark-recapture with occasion and indi-  
1220 vidual effects: abundance estimation through Bayesian model selection in a  
1221 fixed dimensional parameter space," *Journal of agricultural, biological, and*  
1222 *environmental statistics*, 10, 291–305.
- 1223 Fienberg, S. E., Johnson, M. S., and Junker, B. W. (1999), "Classical multilevel  
1224 and Bayesian approaches to population size estimation using multiple lists,"  
1225 *Journal of the Royal Statistical Society of London A*, 163, 383–405.
- 1226 Gardner, B. (2009), "missing," *Missing*, missing.
- 1227 Gardner, B., Reppucci, J., Lucherini, M., and Royle, J. (2010a), "Spatially  
1228 explicit inference for open populations: estimating demographic parameters  
1229 from camera-trap studies," *Ecology*, 91, 3376–3383.
- 1230 Gardner, B., Royle, J. A., Wegan, M. T., Rainbolt, R. E., and Curtis, P. D.  
1231 (2010b), "Estimating black bear density using DNA data from hair snares,"  
1232 *The Journal of Wildlife Management*, 74, 318–325.
- 1233 Gopalaswamy, A. M., Royle, J. A., Delampady, M., Nichols, J. D., Karanth,  
1234 K. U., and Macdonald, D. W. (2012), "Density estimation in tiger popula-  
1235 tions: combining information for strong inference," *Ecology*.
- 1236 Huggins, R. M. (1989), "On the statistical analysis of capture experiments,"  
1237 *Biometrika*, 76, 133.
- 1238 Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. (2008), *Statistical analysis*  
1239 *and modelling of spatial point patterns*, Wiley.
- 1240 Ivan, J. (2012), "Density, demography, and seasonal movements of snowshoe  
1241 hares in central Colorado," Ph.D. thesis, Colorado State University.
- 1242 Johnson (2010), "A Model-Based Approach for Making Ecological Inference  
1243 from Distance Sampling Data," *Biometrics*, 66, 310318.
- 1244 Karanth, K. U. (1995), "Estimating tiger *Panthera tigris* populations from  
1245 camera-trap data using capture–recapture models," *Biological Conservation*,  
1246 71, 333–338.
- 1247 Kendall, W. L., Nichols, J. D., and Hines, J. E. (1997), "Estimating Temporary  
1248 Emigration Using Capture-Recapture Data with Pollock's Robust Design,"  
1249 *Ecology*, 78, 563–578.
- 1250 Kéry, M., Gardner, B., Stoeckle, T., Weber, D., and Royle, J. A. (2010), "Use  
1251 of Spatial Capture-Recapture Modeling and DNA Data to Estimate Densities  
1252 of Elusive Animals," *Conservation Biology*, 25, 356–364.

- 1253 Kéry, M. and Schaub, M. (2011), *Bayesian Population Analysis Using WinBugs*,  
1254 Academic Press.
- 1255 King, R. and Brooks, S. (2001), “On the Bayesian analysis of population size,”  
1256 *Biometrika*, 88, 317–336.
- 1257 King, R., Brooks, S., and Coulson, T. (2008), “Analyzing Complex Capture–  
1258 Recapture Data in the Presence of Individual and Temporal Covariates and  
1259 Model Uncertainty,” *Biometrics*, 64, 1187–1195.
- 1260 Laird, N. M. and Ware, J. H. (1982), “Random-effects models for longitudinal  
1261 data,” *Biometrics*, 963–974.
- 1262 Langtimm, C. A., Dorazio, R. M., Stith, B. M., and Doyle, T. J. (2011), “New  
1263 aerial survey and hierarchical model to estimate manatee abundance,” *The*  
1264 *Journal of Wildlife Management*, 75, 399–412.
- 1265 Link, W. A. (2003), “Missing,” *missing*, missing.
- 1266 Liu and Wu (1999), “Parameter expansion for data augmentation,” *J Am Stat*  
1267 *Assoc*, 94, 1264–1274.
- 1268 MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A., and  
1269 Langtimm, C. A. (2002), “Estimating site occupancy rates when detection  
1270 probabilities are less than one,” *Ecology*, 83, 2248–2255.
- 1271 Mackenzie, D. I. and Royle, J. (2005), “Designing occupancy studies: general  
1272 advice and allocating survey effort,” *Journal of Applied Ecology*, 42, 1105–  
1273 1114.
- 1274 Norris III, J. L. and Pollock, K. H. (1996), “Nonparametric MLE under two  
1275 closed capture-recapture models with heterogeneity,” *Biometrics*, 639–649.
- 1276 Parmenter, R. R., Yates, T. L., Anderson, D. R., Burnham, K. P., Dunnum,  
1277 J. L., Franklin, A. B., Friggens, M. T., Lubow, B. C., Miller, M., Olson, G. S.,  
1278 and Others (2003), “Small-mammal density estimation: A field comparison  
1279 of grid-based vs. web-based density estimators,” *Ecological Monographs*, 73,  
1280 1–26.
- 1281 Pledger, S. (2000), “Unified maximum likelihood estimates for closed capture-  
1282 recapture models using mixtures,” *Biometrics*, 434–442.
- 1283 Royle, J. (2009), “Analysis of capture-recapture models with individual covari-  
1284 ates using data augmentation,” *Biometrics*, 65, 267–274.
- 1285 Royle, J. A. (2006), “Site occupancy models with heterogeneous detection prob-  
1286 abilities,” *Biometrics*, 62, 97–102.
- 1287 — (2008), “Modeling individual effects in the Cormack–Jolly–Seber model: a  
1288 state-space formulation,” *Biometrics*, 64, 364–370.

- (2010), “missing,” *missing*, missing.
- Royle, J. A., Converse, S., and Link, W. (2011a), “Data augmentation for structured populations,” *unpublished*.
- Royle, J. A., Dawson, D. K., and Bates, S. (2004), “Modeling abundance effects in distance sampling,” *Ecology*, 85, 1591–1597.
- Royle, J. A. and Dorazio, R. M. (2008), *Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities*, Academic Press.
- (2010), “Missing,” *Missing*, missing, missing.
- (2011), “Missing,” *Missing*, missing, missing.
- Royle, J. A., Dorazio, R. M., and Link, W. A. (2007), “Analysis of multinomial models with unknown index using data augmentation,” *Journal of Computational and Graphical Statistics*, 16, 67–85.
- Royle, J. A., Magoun, A. J., Gardner, B., Valkenburg, P., and Lowell, R. E. (2011b), “Density estimation in a wolverine population using spatial capture–recapture models,” *The Journal of Wildlife Management*, 75, 604–611.
- Royle, J. A. and Young, K. V. (2008), “A Hierarchical Model For Spatial Capture-Recapture Data,” *Ecology*, 89, 2281–2289.
- Sanathanan, L. (1972), “Estimating the size of a multinomial population,” *The Annals of Mathematical Statistics*, 142–152.
- Schofield, M. and Barker, R. (2008), “A unified capture-recapture framework,” *Journal of agricultural, biological, and environmental statistics*, 13, 458–477.
- Sillett (2011), “Missing,” *Missing*, missing.
- Tanner, M. A. and Wong, W. H. (1987), “The calculation of posterior distributions by data augmentation,” *J Am Stat Assoc*, 82, 528–540.
- Tyre, A. J., Tenhumberg, B., Field, S. A., Niejalke, D., Parris, K., and Possingham, H. P. (2003), “Improving precision and reducing bias in biological surveys: estimating false-negative error rates,” *Ecological Applications*, 13, 1790–1801.
- Wallace, R. B., Gomez, H., Ayala, G., and Espinoza, F. (2003), “Camera trapping for jaguar (*Panthera onca*) in the Tuichi Valley, Bolivia,” *Journal of Neotropical Mammalogy*, 10, 133–139.
- Wegan, M. T. (2008), “Aversive conditioning, population estimation, and habitat preference of black bears (*Ursus americanus*) on Fort Drum Military Installation in northern New York,” Ph.D. thesis, Cornell University, Jan.

- 1324 Wright, J., Barker, R., Schofield, M., Frantz, A., Byrom, A., and Gleeson,  
1325 D. (2009), “Incorporating Genotype Uncertainty into Mark–Recapture-Type  
1326 Models For Estimating Abundance Using DNA Samples,” *Biometrics*, 65,  
1327 833–840.
- 1328 Yang, H. C. and Chao, A. (2005), “Modeling Animals’ Behavioral Response by  
1329 Markov Chain Models for Capture–Recapture Experiments,” *Biometrics*, 61,  
1330 1010–1017.