Andy: Have not checked heading CAPs throughout b/c this is not the proof.

15

MODELS FOR SEARCH-ENCOUNTER DATA

In this chapter we discuss models for search-encounter data. These models are useful in situations where the locations of individuals, say $\mathbf{u}_{ik}$ for individuals i and sample occasions k, are observed directly by searching space (often delineated by a polygon) in some fashion, rather than restricted to fixed trap locations. In all the cases addressed in this chapter, both detection probability and parameters related to movement can be estimated using such models. To formalize this notion a little bit using some of the ideas we've introduced in previous chapter, most of the SCR models we've talked about in the book involve just two components of a hierarchical model, the observation component, which we denote by $[y|\mathbf{s}]$ (e.g., Bernoulli, Poisson, or multinomial), and the process component describing the activity center model $[\mathbf{s}]$, the point process model for the activity centers. The search-encounter models described here involve an additional component for the locations conditional on the activity centers. We write this as follows: The observation model has the form $[y|\mathbf{u}]$, and the process model has two components, a movement model $[\mathbf{u}|\mathbf{s}]$, which describes the individual encounter locations conditional on $\mathbf{s}$, and the point process model $[\mathbf{s}]$. Because we can resolve parameters of the $[\mathbf{u}|\mathbf{s}]$ component, search-encounter models are slightly more complicated, and also more biologically realistic. Conversely, when we have an array of fixed trap locations, the movement process is completely confounded with the encounter process because the list of potential observation locations is prescribed, a priori, independent of any underlying movement process.

A few distinct types of situations exist where search-encounter models come in handy. The prototypical, maybe ideal, situation Royle et al. (2011a) is where we have a single search path through a region of space from which observations are made (just as in the typical distance sampling situation, using a transect). As we walk along the search path, we note the location of each individual that is detected, *and their identity* (this is different from distance sampling in that sense). Alternatively, we could delineate a search area, and conduct a systematic search of that region. An example is that of Royle and Young (2008), which involved a plot search for lizards. They assumed the plot was uniformly searched which justified an assumption of constant encounter probability, p, for all individuals

parentheses

within the plot boundaries. The data set was ≥ 1 location observations for each of a sample of n individuals. The recent paper by Efford (2011a) discussed likelihood analysis of similar models. In the terminology of `secr` such models are referred to as models for *polygon detectors.*

discusses ?

15.1 SEARCH-ENCOUNTER DESIGNS

Before we discuss models for search-encounter data, we'll introduce some types of sampling situations that produce individual location data by searching space. We imagine there are a lot more sampling protocols (and variations) than identified here, but these are some of the standard situations that we have encountered over the last few years in developing applications of SCR models. For our purposes here we recognize 4 basic sampling designs, each of which might have variations due to modification of the basic sampling protocol.

15.1.1 Design 1: Fixed Search Path

A useful class of models arises when we have a fixed search-path or line, or multiple such lines, in some region (Fig. 15.1) from which individual detections are made We assume the survey path is laid out *a priori* in some manner that is done independent of the activity centers of individuals and the collection of data does not affect the lines. The purpose of this assumption, in the models described subsequently, is to allow us to assume that the activity centers are uniformly distributed on the prescribed state-space. Alternatively, explicit models could be entertained to mitigate a density gradient or covariate effects (see Chapt. 11). The situation depicted in Fig. 15.1 shows the search path traversing several delineated polygons, although the polygon boundaries may or may not affect the potential locations of individuals (see below).

A number of variations of this fixed search path situation are possible, and these produce slightly different data structures and corresponding modifications to the model, although we do not address all of these from a technical standpoint here:

Protocol (1a). We know the search path and record the locations of individuals.
Protocol (1b). We record the location of individuals and the location on the search path where we first observed the individual.
Protocol (1c). We record the closest perpendicular distance. This is a typical distance sampling situation, and this is a type of hybrid SCR/distance sampling model.

capital ?

15.1.2 Design 2: Uniform search intensity

In the uniform search intensity model (or just "uniform search"), we have one or more well-defined sample areas (polygons), such as a quadrat or a transect, and we imagine that the area is uniformly searched so that encounter probability is constant for all individuals within the search area. This type of sampling method is often called "area search" in the bird literature (Bibby et al., 1992). Sampling produces locations of individuals within the well-defined boundaries of the sample area. The polygon boundaries defining the sample unit are important because they tell us that $p = 0$ by design outside of the boundary.

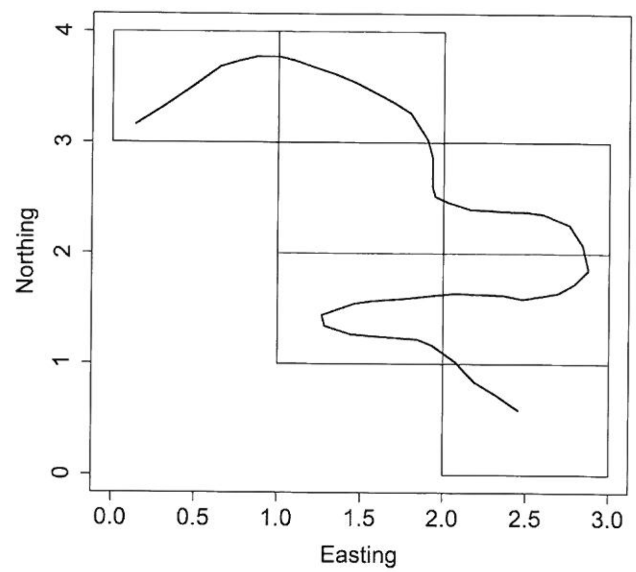

Figure 15.1. A survey line through parts of 7 quadrats in a hypothetical landscape. An observer travels the transect and identifies individuals in the vicinity of the line, recording their identity and location.

Using the example from ~~the~~ Fig. 15.1, but ignoring the survey line through the plot (pretend it doesn't exist), we imagine that each of the identified quadrats is uniformly ? searched, which is to say, we assume that each individual within the boundaries of the → a ? *quadrat* has an equal probability of being detected. In the context of replicate sampling occasions (e.g., on consecutive days), individuals may move on or off of the plot, and so individuals may have different probabilities of being *available* to encounter, based on the closeness of their activity center to the quadrat boundaries. However, given that they're available, the uniform search model assumes they have constant encounter probability.

15.2 A MODEL FOR FIXED SEARCH PATH DATA

In contrast to most of the models described in this book (but see Sec. 9.4), we develop models for encounter probability that depend explicitly on the instantaneous location $\mathbf{u}_{ik}$, for individual i at sample occasion k, say $p_{ik} \equiv p(\mathbf{u}_{ik}) = \Pr(y_{ik} = 1|\mathbf{u}_{ik})$. Note that $\mathbf{u}$ is unobserved for the $y = 0$ observations and thus we cannot analyze the conditional-on-$\mathbf{u}$ likelihood directly. Instead, we regard $\mathbf{u}$ as random effects and assume a model for them, which allows us to handle the problem of missing $\mathbf{u}_{ik}$ values (Sec. 15.4.1). We assume that individuals do not move *during* a sampling occasion or, if they do, the individual is not added to the data set twice.

To develop encounter probability models for this problem we cannot just use the

SCR ← previous models because the "trap" is actually a line or collection of line segments (e.g., Fig. 15.1). Intuitively, $\Pr(y_{ik} = 1|\mathbf{u}_{ik})$ should increase as $\mathbf{u}_{ik}$ comes "close" to the line segments $\mathbf{X}$. It seems reasonable to express closeness by some distance metric $||\mathbf{u}_{ik} - \mathbf{X}||$

as2 ← is the distance between locations $\mathbf{u}_{ik}$ and $\mathbf{X}$, and then assume

something wrong w/ this sentence

$$\text{logit}(p_{ik}) = \alpha_0 + \alpha_1 ||\mathbf{u}_{ik} - \mathbf{X}||.$$

For the case where $\mathbf{X}$ describes a wandering line, some kind of average distance from $\mathbf{u}$ to the line might be reasonable; possible alternatives include the absolute minimum distance or the mean over specific segments of the line (within some distance), etc.... We could also have a model without an explicit distance component, by assuming that individuals within a certain distance from the search path are encountered with equal probability. In this case, we have only a single parameter α_0 but must also specify the distance limit.

15.2.1 Modeling total hazard to encounter

→ capital ?

Because the line $\mathbf{X}$ is not a single point (like a camera trap) we have to somehow describe the total encounter probability induced by the line. A natural approach is to model the total hazard to capture (Borchers and Efford, 2008), which is standard in survival analysis, and also distance sampling (Hayes and Buckland, 1983; Skaug and Schweder, 1999). The individual is detected if encountered at any point along $\mathbf{X}$. Naturally, covariates are modeled as affecting the hazard rate and we think of distance to the line as a covariate acting on the hazard. Let $h(\mathbf{u}_{ik}, \mathbf{x})$ be the hazard of individual i being encountered by sampling at a point $\mathbf{x}$ on occasion t. For example, one possible model assumes, for all points $\mathbf{x} \in \mathbf{X}$,

$$\log(h(\mathbf{u}_{ik}, \mathbf{x})) = \alpha_0 + \alpha_1 * ||\mathbf{u}_{ik} - \mathbf{x}||. \quad (15.2.1)$$

use subscript j (or x) here ?

Additional covariates could be included in the hazard function in the same way as for any model of encounter probability that we've discussed previously. The total hazard to encounter anywhere along the survey path, for an individual located at $\mathbf{u}_{ik}$, say $H(\mathbf{u}_{ik})$, is obtained by integrating over the surveyed line, which we will evaluate numerically by a discrete sum where the hazard is evaluated at the set of points $\mathbf{x}_j$ along the surveyed path: a little awkward

$$H(\mathbf{u}_{ik}) = \exp(\alpha_0) \left\{ \sum_{j=1}^{J} \exp(\alpha_1 * ||\mathbf{u}_{ik} - \mathbf{x}_j||) \right\} \quad (15.2.2)$$

comma

where $\mathbf{x}_j$ is the j^{th} row of $\mathbf{X}$ defining the survey path as a collection of line segments which can be arbitrarily dense, but should be regularly spaced. Then the probability of encounter on a given sampling occasion is

$$p_{ik} \equiv p(\mathbf{u}_{ik}) = 1 - \exp(-H(\mathbf{u}_{ik})). \quad (15.2.3)$$

It is or It's

Its possible that the search path could vary by sampling occasion, say $\mathbf{X}_k$, which can easily be accommodate in the model simply by calculating the total hazard to encounter for each distinct search path.

This is a reasonably intuitive type of encounter probability model in that the probability of encounter is large when an individual's location $\mathbf{u}_{ik}$ is close to the line in the average sense defined by Eq. 15.2.2, and vice versa. Further, consider the case of a single survey point, i.e., $\mathbf{X} \equiv \mathbf{x}$, which we might think of as a camera trap location. In this case note that Eq. (15.2.3) is equivalent to comma?

$$\log(-\log(1 - p_{ik})) = \alpha_0 + \alpha_1 * ||\mathbf{u}_{ik} - \mathbf{x}||$$

comma?

which is to say that distance is a covariate on detection that is linear on the complementary log-log scale, which is similar to the "trap-specific" encounter probability of our Bernoulli encounter probability model (see Chapt. 5). The difference is that, here, the relevant distance is between the "trap" (i.e. the survey lines) and the individual's present location, $\mathbf{u}_{ik}$, which is observable. On the other hand, in the context of camera traps, the distance is that between the trap and a latent variable, $\mathbf{s}_i$, representing an individual's home range or activity center which is not observed. comma

A key assumption of this formulation of the model is that encounters at each point along the line, $\mathbf{x}_j$, are independent of each other point. Then, the event that an individual is encountered *at all* is the complement of the event that it is not encountered *anywhere* along the line (Hayes and Buckland, 1983). In this case, the probability of not being encountered at trap j is: $1 - p(\mathbf{u}_{ik}, \mathbf{x}_j) = \exp(-h(\mathbf{u}_{ik}, \mathbf{x}_j))$ and so the probability that an individual is not encountered at all is $\prod_j \exp(-h(\mathbf{u}_{ik}, \mathbf{x}_j))$. The encounter probability is therefore the complement of this, which is precisely the expression given by Eq. 15.2.3.

use cap. X for line or say "point along the line" point

Any model for encounter probability can be converted to a hazard model so that encounter probability based on total hazard can be derived. We introduced this model above:

$$\log(h(\mathbf{u}_{ik}, \mathbf{x})) = \alpha_0 + \alpha_1 * ||\mathbf{u}_{ik} - \mathbf{x}||.$$

which is usually called the Gompertz hazard function in survival analysis, and it is most often written as $h(t) = a \exp(b * t)$ in which case $\log(h(t)) = \log(a) + b * t$. In the context

comma? what is b?

of survival analysis, t is "time" whereas, in SCR models, we model hazard as a function of distance. The Gaussian model has a squared-distance term:

$$\log(h(\mathbf{u}_{ik}, \mathbf{x})) = \alpha_0 + \alpha_1 * ||\mathbf{u}_{ik} - \mathbf{x}||^2.$$

use subscript j on x 2

Borchers and Efford (2008) use this model:

$$h(\mathbf{u}_{ik}, \mathbf{x}) = -\log(1 - \text{expit}(\alpha_0)\exp(\alpha_1 * ||\mathbf{u}_{ik} - \mathbf{x}||^2))$$

which produces a normal kernel model for *probability of detection* at the point level. i.e., $\Pr(y = 1) = 1 - \exp(-h) = h_0 \exp(\alpha_1 * ||\mathbf{u}_{ik} - \mathbf{x}||^2)$ where $h_0 = \text{logit}^{-1}(\alpha_0)$. Another model is:

$$\log(h(\mathbf{u}_{ik}, \mathbf{x})) = \alpha_0 + \alpha_1 * ||\mathbf{u}_{ik} - \mathbf{x}||$$

which is a Weibull hazard function.

15.2.2 Modeling movement outcomes

We have so far described the model for the encounter data in a manner that is conditional on the locations $\mathbf{u}_{ik}$, some of which are unobserved. Naturally, we should specify a model for these latent variables – i.e., a movement model – so that we could either do a Bayesian analysis by MCMC (Royle and Young, 2008; Royle et al., 2011a) or compute the marginal likelihood (Efford, 2011a). To develop such a model, we adopt what is now customary in SCR models – we assume that individuals are characterized by a latent variable, $\mathbf{s}_i$, which

their?

represents the activity center. This leads to some natural models for the movement outcomes $\mathbf{u}_{ik}$ conditional on the activity center $\mathbf{s}_i$. Royle and Young (2008) used a bivariate normal model:

$$\mathbf{u}_{ik}|\mathbf{s}_i \sim \text{BVN}(\mathbf{s}_i, \sigma^2_{move}\mathbf{I}),$$

where $\mathbf{I}$ is the 2×2 identity matrix. We consider alternatives below. This is a primitive model of individual movements about their home range but we believe it will be adequate in many capture-recapture studies which are often limited by sparse data.

G. White begs to differ... :)

We adopt our default assumption for the activity centers $\mathbf{s}$:

$$\mathbf{s}_i \sim \text{Uniform}(\mathcal{S}); \;\; i = 1, 2, \ldots, N.$$

The usual considerations apply in specifying the state-space $\mathcal{S}$ – either choose a large rectangle, or prescribe a habitat mask to restrict the potential locations of $\mathbf{s}$.

15.2.3 Simulation and analysis in JAGS

Here we will simulate a sample data set that goes with the situation described in Fig. 15.1 and then analyze the data in **JAGS**. We begin by defining the state-space containing all of the grid cells in the rectangle $[-1, 4] \times [-1, 5]$, which contains 30 1×1 cells. The survey line in Fig. 15.1 traverses 7 of those 1×1 boxes. We define the total population to be 4 individuals per grid cell ~~(1 × 1)~~. To set this up in **R**, we do this:

grid cells

```
> xlim <- c(-1, 4)
> ylim <- c(-1, 5)
> perbox <- 4
> N <- 30*perbox   # Total of 30 1x1 quadrats
```

The line in Fig. 15.1 is an irregular mesh of points obtained by an imperfect manual point-and-clicking operation, which mimics the way in which GPS points come to us. In order to apply our model we need a regular mesh of points. We can obtain a regular mesh of points from the irregular mesh by using some functions in the packages **rgeos** and **sp**, especially the function `sample.Line`, which produces a set of equally-spaced points along a line. The **R** commands are as follows (the complete script is given in the function `snakeline`):

scrbook package

do we cite authors of packages? only the first time? YES! Add ref. for rgeos

```
> library(rgeos)
> library(sp)
> line1 <- source("line1.R")

> line1 <- as.matrix(cbind(line1$value$x,line1$value$y))
> points <- SpatialPoints(line1)

> sLine <- Line(points)
> regpoints <- sample.Line(sLine,250,type="regular")  # Key step!
```

Next, we set a random number seed, simulate activity centers and set some model parameters required to simulate encounter history data. In the following commands you can see where the regular mesh representation of the sample line is extracted from the `regpoints` object which we just created:

```
> set.seed(2014)
> sx <- runif(N,xlim[1],xlim[2])
> sy <- runif(N,ylim[1],ylim[2])

> sigma.move <- .35
> sigma <-.4
> alpha0 <- .8
> alpha1 <- 1/(2*(sigma^2))
> X <- regpoints@coords
> J <- nrow(X)
```

comma

Next we're going to simulate data which we do in 2 steps: For each individual in the population and for each of K sample occasions, we simulate the location of the individual as a bivariate normal random variable with mean $\mathbf{s}_i$ and $\sigma_{move} = 0.35$. Next, we compute the encounter probability model using Eq. 15.2.3, with the bivariate normal hazard model, and then retain the data objects corresponding to individuals that get captured at least once. All of this goes according to the following commands:

```
> K <- 10    ##  Sample occasions = 10
> U <- array(NA,dim=c(N,K,2))  ## Array to hold locations
> y <- pmat <- matrix(NA,nrow=N,ncol=K)  ## Initialize
> for(i in 1:N){
+   for(k in 1:K){
+     U[i,k,] <- c(rnorm(1,sx[i],sigma.move),rnorm(1,sy[i],sigma.move))
+     dvec <- sqrt( ( U[i,k,1] - X[,1])^2 + (U[i,k,2] - X[,2])^2  )
```

```
+       loghaz <- alpha0 - alpha1*dvec*dvec
+       H <- sum(exp(loghaz))
+       pmat[i,k] <- 1-exp(-H)
+       y[i,k] <- rbinom(1,1,pmat[i,k])
>     }
>  }
> Ux <- U[,,1]
> Uy <- U[,,2]
> Ux[y==0] <- NA
> Uy[y==0] <- NA
```

In the commands shown above, we define matrices, Ux and Uy, that hold the observed locations of individuals during each occasion. Note that, if an individual is *not* captured, we set the value to NA. We pass these partially observed objects to **JAGS** to fit the model.

Finally, we do the data augmentation and we make up some starting values for the location coordinates that are missing. For these, we cheat a little bit (for convenience and hopefully to improve the efficiency of the MCMC for the simulated data sets) and use the actual activity center values. In practice, we might think about using the average of the observed locations.

```
> ncap <- apply(y,1,sum)
> y <- y[ncap>0,]
> Ux <- Ux[ncap>0,]
> Uy <- Uy[ncap>0,]

> M <- 200
> nind <- nrow(y)
> y <- rbind(y,matrix(0,nrow=(M-nrow(y)),ncol=ncol(y)))
> Namat <- matrix(NA,nrow=(M-nind),ncol=ncol(y))
> Ux <- rbind(Ux,Namat)
> Uy <- rbind(Uy,Namat)
> S <- cbind(runif(M,xlim[1],xlim[2]),runif(M,ylim[1],ylim[2]))
> for(i in 1:nind){
+      S[i,] <- c( mean(Ux[i,],na.rm=TRUE),mean(Uy[i,],na.rm=TRUE))
> }
> Ux.st <- Ux
> Uy.st <- Uy
> for(i in 1:M){
+      Ux.st[i,!is.na(Ux[i,])]<-NA
+      Uy.st[i,!is.na(Uy[i,])]<-NA
+      Ux.st[i,is.na(Ux[i,])]<-S[i,1]
+      Uy.st[i,is.na(Uy[i,])]<-S[i,2]
+ }
```

The **BUGS** model specification is shown in Panel 15.1, although we neglect the standard steps showing how to bundle the data, inits, and farm all of this stuff out to **JAGS** (see the help file for snakeline for the complete script). Simulating the data as described above, and fitting the model in Panel 15.1 produces the results in Table 15.1.

```
model {

 alpha0~dunif(-25,25)              # Priors distributions
 alpha1~dunif(0,25)
 lsigma~dunif(-5,5)
 sigma.move<-exp(lsigma)
 tau<-1/(sigma.move*sigma.move)
 psi~dunif(0,1)

 for(i in 1:M){ # Loop over individuals
   z[i]~dbern(psi)
   s[i,1]~dunif(xlim[1],xlim[2])    # Activity center model
   s[i,2]~dunif(ylim[1],ylim[2])
   for(k in 1:K){                   # Loop over sample occasions
      ux[i,k] ~ dnorm(s[i,1],tau)  # Movement outcome model
      uy[i,k] ~ dnorm(s[i,2],tau)
      for(j in 1:J){ # Loop over each point defining line segments
        d[i,k,j]<-  pow(pow(ux[i,k]-X[j,1],2) + pow(uy[i,k]-X[j,2],2),0.5)
        h[i,k,j]<-exp(alpha0-alpha1*d[i,k,j]*d[i,k,j])
     }
    H[i,k]<-sum(h[i,k,1:J])         # Total hazard H
    p[i,k]<- z[i]*(1-exp(-H[i,k]))
    y[i,k] ~ dbern(p[i,k])
  }
}
 # Population size is a derived quantity
 N<-sum(z[])
}
```

Panel 15.1: **BUGS** model specification for the fixed search path model, based on that from Royle et al. (2011a). See the help file ?snakeline for the **R** code to simulate data and fit this model.

what?

Table 15.1. Posterior summary statistics for the simulated fixed search path data. These are based on 3 chains, and a total of 9000 posterior samples. The data generating parameter values were $N = 100$, $\sigma_{move} = 0.35$, $\sigma = 0.4$, and $\alpha_0 = 0.8$. The parameter $\alpha_1 = 1/(2\sigma^2)$.

order of params?

Parameter	Mean	SD	2.5%	50%	97.5%	Rhat
N	117.626	5.675	107.000	117.000	129.000	1.015
α_0	1.305	0.494	0.425	1.280	2.387	1.009
α_1	3.806	0.423	3.050	3.777	4.733	1.008
σ_{move}	0.347	0.008	0.332	0.347	0.364	1.023
σ	0.364	0.020	0.325	0.364	0.405	1.008
ψ	0.587	0.044	0.501	0.588	0.673	1.006

15.2.4 Hard plot boundaries

CAP?

The previous development assumed that locations of individuals can be observed anywhere in the state-space, determined only by the encounter probability model as a function of distance from the search path. However, in many situations, we might delineate a plot which restricts where individuals might be observed (as in the situation considered by Royle and Young (2008)). For such cases we truncate the encounter probability function at the plot boundary, according to:

$$p(\mathbf{u}_{ik}) = (1 - \exp(-H(\mathbf{u}_{ik})))\mathrm{I}(\mathbf{u}_{ik} \in \mathcal{X}) \tag{15.2.4}$$

where $\mathcal{X}$ is the surveyed polygon and the indicator function $\mathrm{I}(\mathbf{u}_{ik} \in \mathcal{X}) = 1$ if $\mathbf{u}_{ik} \in \mathcal{X}$ and 0 otherwise. That is, the probability of encounter is identically 0 if an individual is located *outside* the plot at sample period t. We demonstrated how to do this in the **BUGS** language below for a model of uniform search intensity (area-search model).

15.2.5 Analysis of other protocols

CAP?

no comma?

In the situation elaborated on above (what we called "Protocol 1a"), the sample path is used to locate individuals and, whether or not an individual is encountered, is a function of the total hazard to encounter along the whole line. We think there are a number of variations of this basic design that might arise in practice. A slight variation (what we called "Protocol 1b") is based on recording location of individuals and also the location on the transect where we observed the individual. The probability of encounter is the probability of encounter prior to the point on the line where the detection takes place (Skaug and Schweder, 1999). This is exactly a distance-sampling observation model, but with an additional hierarchical structure that the describes the individual locations about their activity centers. There are no additional novel considerations in analysis of this situation compared to Protocol 1a, and so we have not given it explicit consideration here. Similarly, "Protocol 1c" is a slight variation of this – instead of recording the point on the line where the individual was first detected, we use~~, instead,~~ the point on the line that has the shortest perpendicular distance. This is a classical distance sampling observation model, and it represents an intentional misspecification of the model but it seems that the effect of this is relatively minor, or, otherwise, we imagine people wouldn't do it.

15.3 UNSTRUCTURED SPATIAL SURVEYS

no comma?

A common situation in practice is that in which sampling produces a survey path, but the path was not laid out *a priori* but, rather evolves opportunistically during the course of sampling, a situation we'll call an unstructured spatial survey (Thompson et al., 2012; Russell et al., 2012). We imagine that the survey path evolves in response to information about animal presence, which could be both the number of unique individuals or the amount of sign in the local search area. The motivating problem has to do with area searches using dog teams, in which the dogs usually wander around hunting scat, and their search path is based on how they perceive the environment and what they're smelling. This violates the main assumptions that the line is placed a priori, independent of density and unrelated to detectability.

The analysis framework implemented by Thompson et al. (2012) and Russell et al. (2012) is based on a heuristic justification wherein the sampling of space is imagined to have been grid-structured, with grid cells that are large enough so that dogs are not influenced by scat or sign beyond the specific cell being searched. Then, we assume the dog applies a consistent search strategy to each cell so that that resulting cell-level detections can be regarded as independent Bernoulli trials with probability p_{ij} depending on the distance $||\mathbf{x}_j - \mathbf{s}_i||$ between the grid cell with center $\mathbf{x}_j$, and individual with activity center $\mathbf{s}_i$ and the amount of search effort (or length of the search route) within a cell. In other words, we use an ordinary SCR type of model but treating the center point of each cell as an effective "trap". The deficiency with this approach is that some of the "sub-grid" resolution information about movement is lost, so we probably lose precision about any parameters of the movement model when the cells are large relative to a typical home range size. We discuss a couple of examples below.

15.3.1 Mountain lions in Montana

CAP?

Russell et al. (2012) analyzed mountain lion (*Puma concolor*) encounter history data to assess the status of mountain lions in the Blackfoot Mountains of Montana. The data collection was based on opportunistic searching by hunters with dogs, who tree the lion (Fig. 15.2). Tissue is extracted with a biopsy dart and analyzed in the lab for individual identity. They used 5 km × 5 km grid cells for binning the encounters, and the length of the search path in each grid cell as a covariate of effort (C_j) that each grid cell was searched. The model is the Gaussian hazard model with baseline encounter probability that depended on sex and effort in each grid cell, on the log scale:

The authors

awkward, maybe just "for each grid cell"

$$\log(\lambda_{0,ij}) = \alpha_0 + \alpha_2 \log(C_j) + \alpha_3 \mathrm{Sex}_i$$

Note for grid cells that were not searched, $C_j = 0$ and, for those, the constraint $\lambda_{0,ij} = 0$ was imposed so that the probability of encounter was identically 0.

One problem encountered by Russell et al. (2012) in their analysis is the possibility of dependence in encounters because of group structure in the data (usually, juveniles in association with their mother). In this situation, in addition to dependence of encounter, multiple individuals have effectively the same activity center, thus violating a number of assumptions related to the ordinary SCR model. To resolve this problem, the authors made some assumptions about group association and fitted models where group served as the functional individual.

Figure 15.2. Mountain lion. Run! *Photo credit: Bob Wiesner.*

15.3.2 Sierra National Forest Fisher Study

This is a little strange bc there is actually less text about this example. Maybe change sentence?

Here we consider a more detailed example and provide the data and **R** script for this analysis. The data come from an analysis of individual encounter histories of the fisher (*Martes pennanti*) by Thompson et al. (2012). The survey area was divided into 15 approximately 1,400 ha hexagons (Fig. 15.3), which is roughly the size of a female fisher's home range, and each hexagon was surveyed 3 times by sniffer dog teams searching space for scat. The dogs were given considerably latitude to determine their route. Thus, the search path ~~is~~ was not laid out a priori but rather evolves opportunistically, based on what the dog senses at a local scale. The authors divided the region into 1 km grid cells (also shown in Fig. 15.3).

* where ?

considerable

I think when this info is basically an adjective, like here, you use a dash between the number and the unit: "a 1-km grid cell"

We provide the data from this study in the `scrbook` package, and it can be loaded with the command `data(fisher)`. The **R** script `SCRfisher` produces the posterior summary statistics shown in Table 15.2. One thing is relatively poor mixing of the Markov chains here due to sparse data and a fairly long run is probably necessary.

issue?

15.4 DESIGN 2: UNIFORM SEARCH INTENSITY

A special case of a search-encounter type of model arises when it is possible to subject a quadrat (or quadrats) to a uniform search intensity. This could be interpreted as an exhaustive search, or perhaps just a thorough systematic search of the available habitat. The example considered by Royle and Young (2008) involved searching a 9 ha plot for horned lizards (Fig. 15.4) by a crew of several people. It was felt in that case that complete

Table 15.2. Posterior summary statistics for the fisher study data, based on 30000 posterior samples. Here $\lambda_0 = \exp(\alpha_0)$. This example exhibits relatively poor mixing due to sparse data, and the Rhat statistic should be reduced by obtaining a larger posterior sample.

order params?

Parameter	Mean	SD	2.5%	50%	97.5%	Rhat
N	315.889	230.041	12.000	280.000	738.775	1.133
σ	4.745	2.909	0.163	4.650	9.704	1.020
λ_0	0.003	0.033	0.000	0.000	0.016	1.097
α_1	0.188	0.170	0.005	0.138	0.641	1.002
ψ	0.413	0.300	0.016	0.366	0.964	1.131

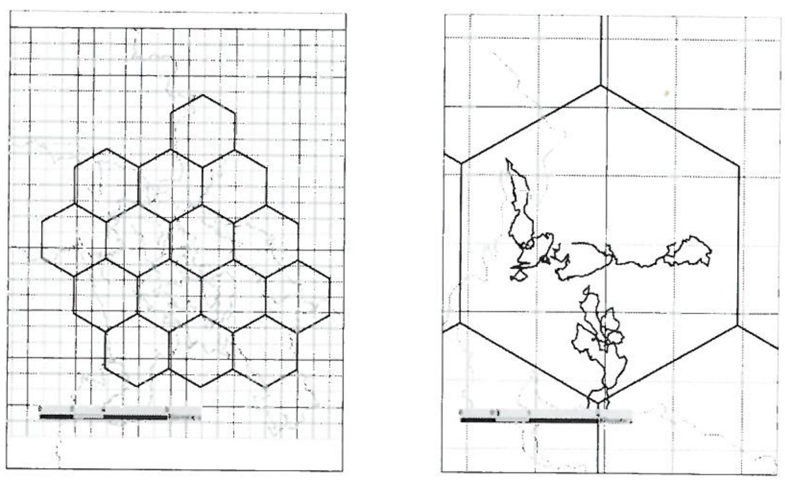

Figure 15.3. Fisher study area showing the gridding system (left panel). The larger hexagons are approximately the size of a typical female home range. The 1-km grid cells define the SCR model grid, where the center point of each one served as a "trap." The right panel shows the GPS trackline of the dog team through one of the ~~grid cells~~. The total length of the trackline was used as a covariate on encounter probability. *Credit: Craig Thompson, U.S. Forest Service*

and systematic (i.e., uniform) coverage of the plot was achieved. In general, however, we think you could have a random sample of the plot and approximate that as a uniform coverage – this is ~~kind of~~ a design-based argument justifying the uniform search intensity model (we haven't simulated this situation, but it would be worth investigating).

* where?

Figure 15.4. A flat-tailed horned lizard showing its typical cryptic appearance in its native environment.* Detection of flat-tailed horned lizards is difficult because they do not run when approached. Instead they shuffle under the sand or press down and remain motionless as shown in the picture. The horns are employed only as a last resort if the camouflage fails. *Photo credit: Kevin and April Young*

It is clear that this uniform search intensity model is a special case of the fixed search path model in the sense that the probability of encounter of an individual is a constant p_0 *if* the individual is located in the polygon $\mathcal{X}$ during sample occasion k, i.e.,

$$p(\mathbf{u}_{ik}) = p_0 \mathrm{I}(\mathbf{u}_{ik} \in \mathcal{X})$$

which resembles Eq. 15.2.4 except replacing the encounter probability function with constant p_0.

Subsequently, we give a simple analysis using simulated data and simple movement models for $\mathbf{u}$, including a bivariate normal model and a random walk. For further examples and analyses, we refer you to Royle and Dorazio (2008), who reanalyzed the lizard data from Royle and Young (2008), and Efford (2011b) and Marques et al. (2011).

15.4.1 Alternative movement models (CAP?

As with the general fixed search path model ("Design 1"), we require a model to describe the movement outcomes $\mathbf{u}_{ik}$. In the analysis of Royle and Young (2008), a simple bivariate

Gaussian movement model was used, in which

BVN

$$\mathbf{u}_{ik}|\mathbf{s}_i \sim \cancel{\text{Normal}}(\mathbf{s}_i, \sigma^2_{move}\mathbf{I}),$$

However, clearly more general versions of the model can be developed. For example, imagine a situation where the successive surveys of a bounded sample polygon are relatively close together in time so that successive locations of individuals are not well-approximated by the Gaussian movement model, which implies independence of locations. Naturally we might consider using an auto-regressive or random-walk type of model in which the successive coordinate locations of individual i behave as follows:

$$u_{1,i,k}|u_{1,i,k-1} \sim \text{Normal}(u_{1,i,k-1}, \sigma^2_{move})$$
$$u_{2,i,k}|u_{2,i,k-1} \sim \text{Normal}(u_{2,i,k-1}, \sigma^2_{move})$$

CAP

here we use the notation u_1 and u_2 for the easting and northing coordinates, respectively. (and, for clarity, we are using commas in the sub-scripting here when we have to refer to time-lags). In addition, we require that the initial locations have a distribution and, for that, we might begin with a simple model such as the uniformity model:

$$\mathbf{u}_{i,1} \sim \text{Uniform}(\mathcal{S}),$$ comma

no dashes

which effectively takes the place of the model for $\mathbf{s}_i$ that we typically use. Under this model, individuals don't have an activity center but, rather, they drift through space more-or-less randomly based just on their previous location. See Ovaskainen (2004) and Ovaskainen et al. (2008) for development and applications of similar movement models in the context of capture-recapture data, and also our discussion of a similar model that might arise in acoustic surveys (Sec. 9.4). We could allow for dependent movements about a central location $\mathbf{s}_i$ using a bivariate auto-regression or similar type of model with parameter ρ, e.g.,

$$\mathbf{u}_{i,k}|\mathbf{s}_i \sim \text{BVN}(\rho * (\mathbf{u}_{i,k-1} - \mathbf{s}_i), \sigma^2_{move}\mathbf{I}).$$

We don't have any direct experience fitting these movement models to real capture-recapture data, but we imagine they should prove effective in applications that yield large sample sizes of individuals and recaptures.

15.4.2 Simulating and fitting uniform search models

The **R** script `uniform_search`, in the `scrbook` package, ~~we~~ provides a script for simulating and fitting search-encounter data using the iid Gaussian model and also the random walk model. The **BUGS** model specification is shown in Panel 15.2 for the random walk situation. We encourage you to adapt this model and the simulation code for the auto-regression movement model. To fit this model to data, we set up the run with **JAGS** using the standard commands. We did not specify starting values for the missing coordinate locations, although we imagine that **JAGS** should perform better if we provide decent starting values, e.g., the last observed location or some other reasonable location. We imagine that resource selection could be parameterized in this movement model as well, perhaps using similar ideas to those described in Chapt. 13.

italics

comma

italics

occasions
model,
locations are

The following script simulates a population of N individuals and their locations at each of 4 ~~times~~ to see if they are in a square [3,13] or not. ~~This simulates a~~ random walk ~~thing~~ so we imagine that the sampling occasions are close together in time. The initial ~~state~~ is assumed to be uniformly distributed on the state-space which, in this case, is the square $[0, 16] \times [0, 16]$. We store the movement outcomes here in a 3-d array `U`, instead of in two separate 2-d arrays (one for each coordinate) as we did above. The **R** commands are as follows:

*Subsequent locations follow a
comma
comma

```
> N <- 100
> nocc <- 4
> Sx <- Sy <- matrix(NA,nrow=N,ncol=nocc)
> sigma.move <- .25

# Simulate initial coordinates on the square:
> Sx[,1] <- runif(N,0,16)
> Sy[,1] <- runif(N,0,16)

> for(t in 2:nyear){
+    Sx[,t] <- rnorm(N,Sx[,t-1],sigma.move)
+    Sy[,t] <- rnorm(N,Sy[,t-1],sigma.move)
+ }

# Now we generate encounter histories on a search rectangle
#     with sides [3,13]:
> Y <- matrix(0,nrow=N,ncol=nyear)
> for(i in 1:N){
+   for(t in 1:nyear){
+     # IF individual is in the sample unit we can capture it:
+     if( Sx[i,t] > 3 & Sx[i,t]< 13 & Sy[i,t]>3 & Sy[i,t]<13 )
+     Y[i,t] <- rbinom(1,1,.5)
+   }
+ }

                     space
# Subset data. If an individualis never captured, cannot have him in our data set
> cap<- apply(Y,1,sum) > 0
> Y <- Y[cap,]
> Sx <- Sx[cap,]
> Sy <- Sy[cap,]

> Sx[Y==0] <- NA
> Sy[Y==0] <- NA

## Data augmentation:
> M <- 200
> Y <- rbind(Y,matrix(0,nrow=(M-nrow(Y)),ncol=nyear))
> Sx <- rbind(Sx,matrix(NA,nrow=(M-nrow(Sx)),ncol=nyear))
> Sy <- rbind(Sy,matrix(NA,nrow=(M-nrow(Sy)),ncol=nyear))
```

```
# Make 3-d array of coordinates "U"
> U <- array(NA,dim=c(M,nyear,2))
> U[,,1] <- Sx
> U[,,2] <- Sy
```

```
model{
psi ~ dunif(0,1)                    # Prior distributions
tau ~ dgamma(.1,.1)
p0 ~ dunif(0,1)
sigma.move <- sqrt(1/tau)

for (i in 1:M){
  z[i] ~ dbern(psi)
  U[i,1,1] ~ dunif(0,16)            # Initial location
  U[i,1,2] ~ dunif(0,16)

   for (k in 2:n.occasions){
      U[i,k,1] ~ dnorm(U[i,k-1,1], tau)
      U[i,k,2] ~ dnorm(U[i,k-1,2], tau)
    }
   for(k in 1:n.occasions){
         # Test whether the actual location is in- or outside the
         #     survey area. Needs to be done for each grid cell
      inside[i,k] <- step(U[i,k,1]-3) * step(13-U[i,k,1]) *
                     step(U[i,k,2]-3) * step(13-U[i,k,2])
      Y[i,k] ~ dbern(mu[i,k])
      mu[i,k] <- p0 * inside[i,k] * z[i]
      }
   }
N <- sum(z[])                       # Population size, derived
}
```

comma

Panel 15.2: **BUGS** model specification for the uniform search intensity model similar to Royle and Young (2008) but with a random walk movement model. help file ?uniform_search in the **R** package scrbook.

Also see the

lower case

15.4.3 Movement and Dispersal in Open Populations

In Chapt. 16 we discuss many aspects of modeling open populations, including some aspects of modeling movement and dispersal and the relevance of SCR models to these comma problems. However, ~~given the introduction of~~ the uniform search model above, ~~this~~ is clearly relevant to modeling movement and dispersal in open populations. In particular, the model described in Panel 15.2 could easily be adapted to an open population by conditioning on the first, and introducing a latent "alive state" with survival parameter ϕ_t. This would be a spatial version of the standard Cormack-Jolly-Seber model (Chapt. 16.3)[1].

?

connection unclear: how are these models linked? need to say s/th about the movement component.

15.5 PARTIAL INFORMATION DESIGNS

The prototype search-encounter (Design 1) and uniform search (Design 2) cases are ideal in the sense that they produce both precise locations of individuals and also a precise characterization of the manner in which individuals are encountered by sampling space. We have seen a number of studies that, in an ideal world, would have generated data consistent with one of these situations but, for some practical reason or other ~~reason~~, partial or no spatial information about the search area or the locations of individuals was collected (or retained), and so the models described above could not be used. We imagine (indeed, have encountered) at least 3 distinct situations:

(a) The search path is not recorded, but locations of individuals are recorded.
(b) The search path is recorded, but locations of individuals are not.
(c) The search path is not recorded, and the locations are not recorded, just raw summaries for prescribed areas or polygons.

For analysis of these search-encounter designs with partial information, we see a number of options of varying levels of formality, depending on the situation (and these are largely untested). For (a) You could always assume uniform search intensity, which might be reasonable if the plots were randomly searched. Otherwise, ~~its~~ validity would depend on the precise manner in which the search activity occurred. For (b) or (c), we could adopt the approach we took in the fisher analysis above, and map the locations to the center of each plot, thinking of the plot as an effective trap, and using the search path length as a covariate. A 4th case with even less information is that in which we don't record individual identity at all. Instead, we just have total count frequencies in each plot. This model is precisely the one considered by (Chandler and Royle, 2013) and this is the focus of Chapt. 18.

lower case

the validity of this assumption

15.6 SUMMARY AND OUTLOOK

The generation of spatial encounter history data in ecological studies is widespread. While such data have historically been obtained mostly by the use of arrays of fixed traps (catch traps, camera traps, etc..), in this chapter we showed that SCR models are equally relevant to a large class of "search-encounter" problems which are based on organized or opportunistic searches of spatial areas. Standard examples include "area searches" in bird population studies, use of detector dogs to obtain scat samples, from which DNA can be obtained to determine individual identity, or sampling along a fixed search path (or transects) by observers noting the locations of detected individuals (this is common in

comma

[1]Some work related to this is currently being carried out by our colleagues Torbjørn Ergon and Michael Schaub.

sampling for reptiles and amphibians). The latter situation closely resembles distance sampling but, with repeated observations of the same individual (on multiple occasions), it has a distinct capture-recapture element to it. In a sense, the fixed search path models are hybrid SCR-DS models.

Many models for search-encounter data have three elements in common. They contain: (1) a model for encounter conditional on locations of individuals; (2) a model that describes how these observable animal locations are distributed in space ~~about their activity centers~~; could be Markovian and (3) a model for the distribution of activity centers.* We interpret the 2nd model component as an explicit movement model, and the existence of this component is distinct from most of the other models considered in this book. One of the key conceptual points is that, with these search-encounter types of designs, the locations of observations are *not* biased by the locations of traps, but, rather, locations of individuals can occur anywhere within search plots or quadrats, or in the vicinity of a transect or search path. Because we can obtain direct observations of location – outcomes of movement – for individuals, it is possible to resolve explicit models of movement from search-encounter data. We considered the simple case of the independent bivariate normal movement model, and also a random walk type model,* which can easily be fitted in the **BUGS** engines. We imagine that much more general movement models can be fitted, although we have had limited opportunities to pursue this and in most practical capture-recapture studies, we will probably be limited by sparse data in the complexity of the movement models that could be considered.

* or initial locations

comma

* both of

Search-encounter sampling is fairly common, although we think that many people don't realize that it can produce encounter history data that is amenable to the development of formal models for density, movement and space usage. We believe that these protocols will become more appealing as methods for formal analysis of the resulting encounter history data become more widely known. At the same time, search-encounter models will increase in relevance in future studies of animal populations because so many new methods of obtaining encounter history data can be be based on DNA extracted from animal tissue or scat, which is easy to obtain by searching space opportunistically. In addition, as the cost of obtaining individual identity from scat or tissue decreases, its widespread collection and use in capture-recapture models can only increase.

is that right? what does that refer to? not sure...