

Latent abundance mixture models for mammals and birds in La Gran Sabana, Venezuela

Ferrer-Paris, J.R.; Stachowicz, I; Sanchez-Mercado, A.

January 20, 2021

We want to test the response of mammal and bird species to the proximity of “conucos” in an area of the Gran Sabana in Venezuela using data from a pre-existing camera trap survey.

First we describe the general workflow for fitting the model for one species (*Dasyprocta leporina*) in four steps: (1) setup of analysis, (2) model fitting, (3) GoF, model criticism, and model selection, and (4) analysis of results.

Then we run this process for all target species in a batch and present the summary of result for all of them at the end of the document.

Model for *Dasyprocta leporina*

Setup of the analysis

We load up the R packages needed for the analysis and load the Rdata file with all the data.

```
# load packages
require(unmarked)
require(AICcmodavg)
require(chron)
require(raster)
require(cluster)
require(MuMIn)
require(reshape2)
require(dplyr)
require(ggplot2)
require(RColorBrewer)
# set working directory
setwd(work.dir)
# load Rdata file from repository
GIS.data <- sprintf("%s/Rdata/GIS.rda",script.dir)
load(GIS.data)
```

Sampling design and camera trap location

We used data from a previous camera trap survey conducted between September 2015 – April 2016. The original sampling design was developed to optimize covering habitat diversity in order to evaluate how mammal species richness is related to habitat types, and is described in detail in Stachowicz et al. (2020), but we provide here a brief summary of the initial setting and how we adapted data a posteriori for our analysis.

Sampling design comprised six 50km^2 blocks within the study area (B01 – B06) each one subdivided into 25 sampling units of 2km^2 .

The sampling universe included $N = 150$ sampling units, but only 30 cameras were available, thus sampling was divided into three periods of 60-days each, and in each period a two-levels stratified random sampling was used to select 30 sampling units (five in each block) for camera deployment. This stratification ensures a balanced representation of sampling units with different coverage of habitat types and fragmentation in each block during each period. As a side effect of this, some sampling units with unique values within each block (for example cells with high tree cover within a block dominated by savanna) were selected for sampling in two or three periods and those cameras were neither relocated nor replaced.

We filter the camera data and show their location in relation with the blocks 1 to 6 located in the eastern part of the Gran Sabana on the border of the Canaima National Park, circle size and colour is proportional to the time the camera was active (duration):

```
camaras %>% filter(bloque %in% sprintf("B%02i",1:6)) %>%
  mutate(bloque=droplevels(bloque),
         fecha1 = chron(dates=as.character(fecha.act), times=as.character(hora.act),
                        format = c(dates = "y-m-d", times = "h:m:s")),
         fecha2 = chron(dates=as.character(fecha.desact.real),
                        times=as.character(hora.desact.real),
                        format = c(dates = "y-m-d", times = "h:m:s")),
         cdg = as.character(ID.original)) %>%
  group_by(cdg) %>%
  summarise(lat=mean(lat), lon=mean(lon), bloque=unique(bloque),
            hunting=unique(factor(caza.celda>0)), grp=unique(grp), H=mean(H), h=mean(h),
            tree_0500m=mean(tree_0500m), tree_1000m=mean(tree_1000m),
            tree_2500m=mean(tree_2500m), tree_5000m=mean(tree_5000m),
            drios=mean(drios), bsq=mean(buf.fragmen), ndvi=mean(ndvi.mu),
            fecha1=min(fecha1), fecha2=max(fecha2)) %>%
  mutate(duration=as.numeric(fecha2-fecha1)) ->
  cam.data
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

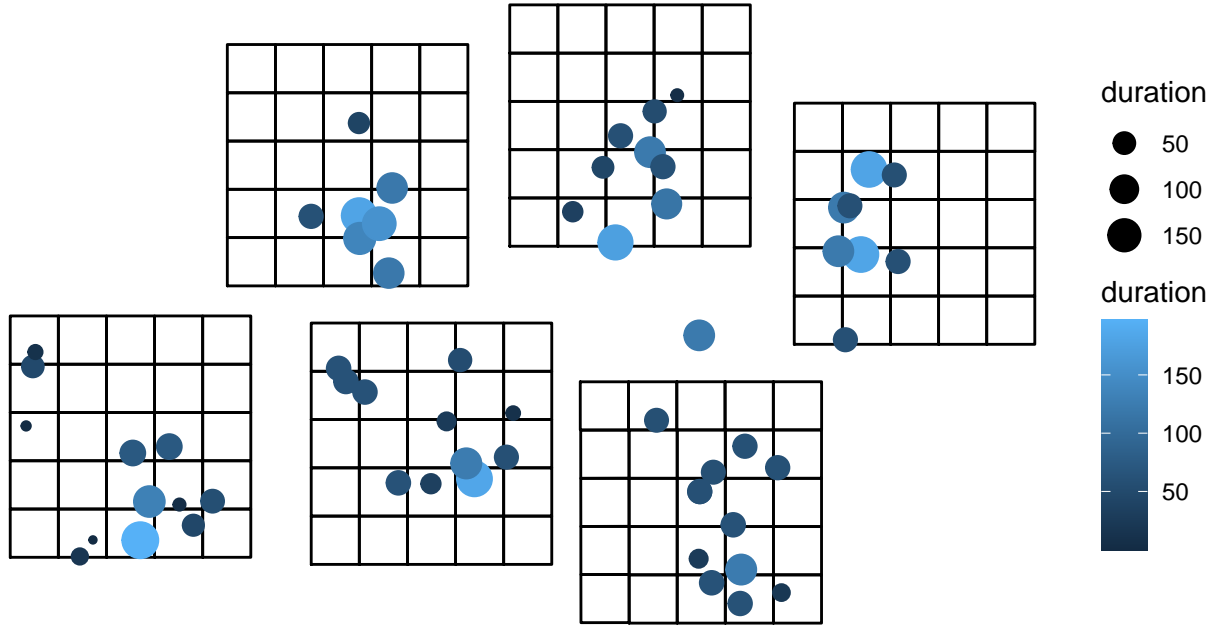
We map the location of cameras in relation to the blocks

```
sampling.design <- subset(grd,cuadrado %in% 1:6)

map <- ggplot() +
  geom_polygon(data = sampling.design,
              aes(x = long, y = lat, group = group), colour = "black", fill = NA)

## Regions defined for each Polygons

map + theme_void() + coord_equal() +
  geom_point(data=cam.data,
            mapping=aes(x=lon,y=lat,size=duration,colour=duration))
```



Vegetation / habitat types

Blocks were selected to represent landscapes with different values of total forest cover (h) and landscape fragmentation index (H ; Stachowicz et al. 2020) :

```
cam.data %>% select(bloque,H,h) %>% unique()
```

```
## # A tibble: 6 x 3
##   bloque      H      h
##   <fct> <dbl> <dbl>
## 1 B04     1.54  39.4
## 2 B02     1.64   5.4
## 3 B06     1.61  20.3
## 4 B05     1.5   53
## 5 B01     1.69   0.4
## 6 B03     1.62  82.5
```

Vegetation is dominated by scrub (*Clusia* spp. and *Gongylolepis* spp.), broadleaf grassland and savannas of *Axonopus* spp. and patches of gallery forest around the rivers, and evergreen montane surrounding the Ilú and Tramén tepuis massif. Sampling units were selected to represent the different habitat types present in each block:

```
cam.data %>% select(bloque,grp) %>% table()
```

```
##           grp
## bloque savanna shrub forest
## B01      12      0      0
## B02       6      3      0
## B03       0      0      8
## B04       1      1      5
## B05       3      9      1
## B06       4      5      2
```

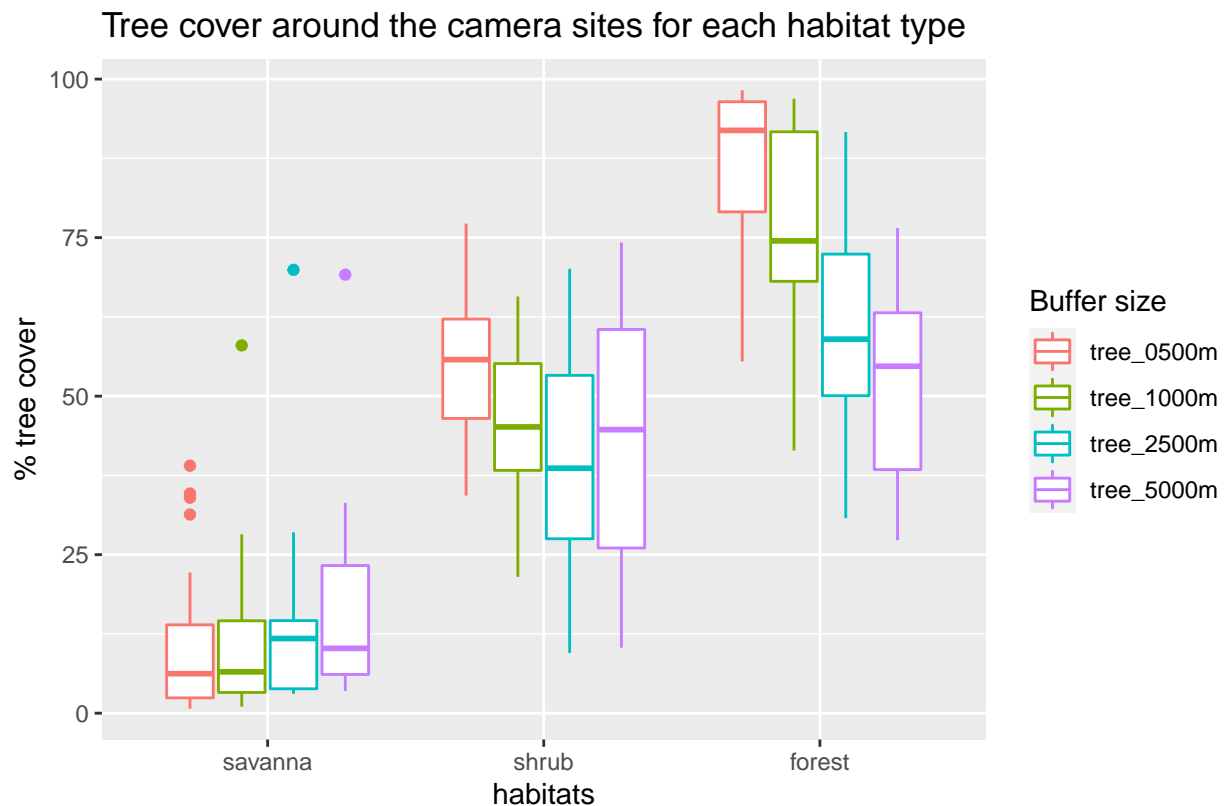
Previous analysis suggested that most mammal species in the study area are associated with forest habitat and few are associated with shrubs or savanna (Stachowicz et al. 2020). We used mean tree cover (in

percentage) as a quantitative variable correlated with these habitat types and consistent with metrics used for the sampling design. Mean tree cover was calculated from remote sensing products (Hansen et al. 2003) using a 1 km buffer around the camera location, the variable has a bimodal distribution with a lower mode at 10-20% corresponding with the savanna, a higher mode at 70-80% corresponding with forest and intermediate values roughly corresponding with the less common shrub habitat.

These groups have different values of tree cover and vegetation indices estimated from remote sensors

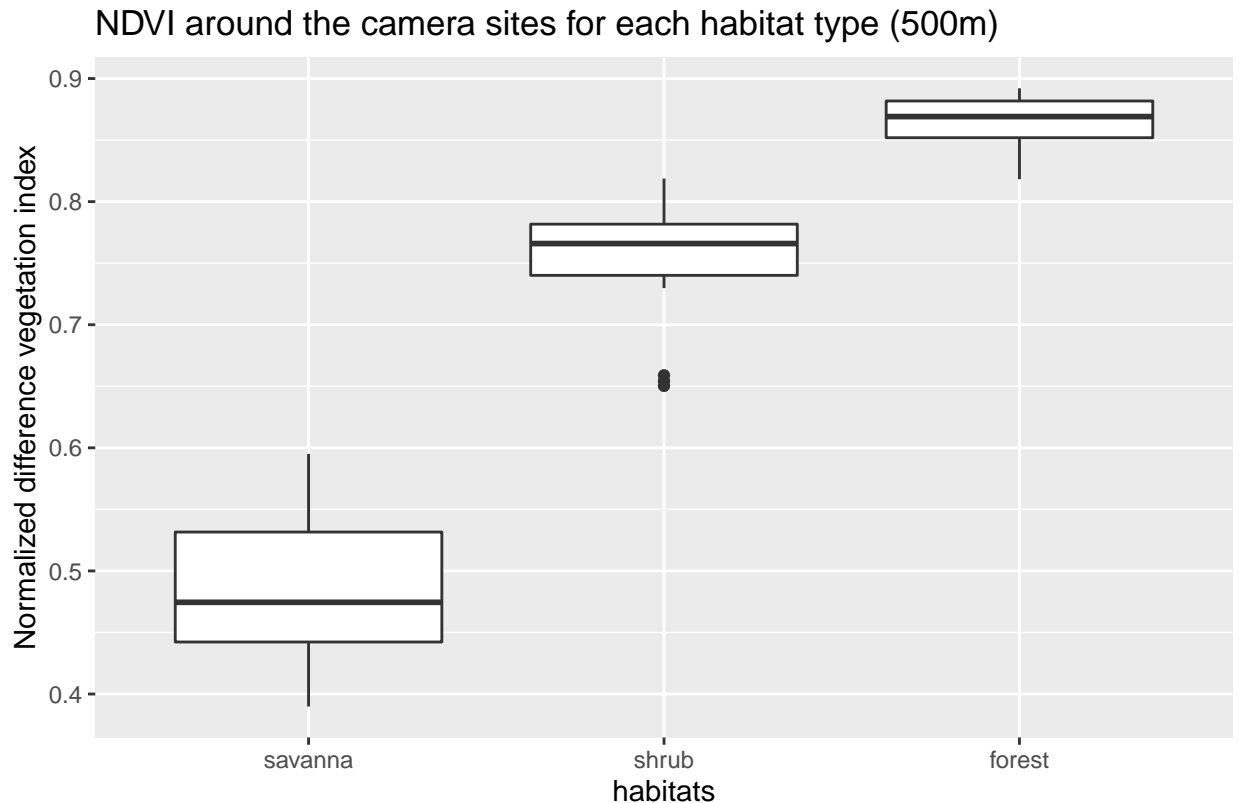
```
dat.m <- melt(cam.data, id.vars='grp', measure.vars=c('tree_0500m', 'tree_1000m', 'tree_2500m', 'tree_5000m'))
```

```
ggplot(dat.m) +
  geom_boxplot(aes(x=grp, y=value, color=variable)) +
  labs(title="Tree cover around the camera sites for each habitat type") +
  labs(y='% tree cover', x="habitats", color='Buffer size',
       caption="Tree cover from Hansen et al. (2013)")
```



Tree cover from Hansen et al. (2013)

```
ggplot(cam.data, aes(x=grp, y=ndvi)) +
  geom_boxplot(notch=F) + # or notch=T
  labs(title="NDVI around the camera sites for each habitat type (500m)") +
  labs(y='Normalized difference vegetation index', x="habitats",
       caption="NDVI derived from MODIS")
```



NDVI derived from MODIS

These two measures of vegetation are significantly correlated with landscape metrics for the blocks:

```
cam.data %>% select(bsq,ndvi,H,h,tree_0500m,tree_1000m,tree_2500m,tree_5000m,drios) %>% cor()
```

```
##          bsq      ndvi      H      h tree_0500m tree_1000m
## bsq      1.0000000  0.8309999 -0.3761795  0.6642176  0.8862003  0.8329604
## ndvi      0.8309999  1.0000000 -0.5309255  0.7490655  0.9607293  0.9088306
## H       -0.3761795 -0.5309255  1.0000000 -0.6004536 -0.4848918 -0.4907362
## h        0.6642176  0.7490655 -0.6004536  1.0000000  0.7434643  0.7838421
## tree_0500m 0.8862003  0.9607293 -0.4848918  0.7434643  1.0000000  0.9574364
## tree_1000m 0.8329604  0.9088306 -0.4907362  0.7838421  0.9574364  1.0000000
## tree_2500m 0.6915607  0.8282899 -0.6210522  0.8847185  0.8436329  0.9137586
## tree_5000m 0.5959857  0.7634414 -0.7159879  0.9030378  0.7555087  0.8130623
## drios     -0.1488141 -0.2985787  0.4527932 -0.1444651 -0.2375248 -0.2297796
##          tree_2500m tree_5000m      drios
## bsq      0.6915607  0.5959857 -0.1488141
## ndvi      0.8282899  0.7634414 -0.2985787
## H       -0.6210522 -0.7159879  0.4527932
## h        0.8847185  0.9030378 -0.1444651
## tree_0500m 0.8436329  0.7555087 -0.2375248
## tree_1000m 0.9137586  0.8130623 -0.2297796
## tree_2500m 1.0000000  0.9545729 -0.2311191
## tree_5000m 0.9545729  1.0000000 -0.2137828
## drios     -0.2311191 -0.2137828  1.0000000
```

```
with(cam.data,cor.test(bsq,H))
```

```
##
```

```
## Pearson's product-moment correlation
##
## data:  bsq and H
## t = -3.092, df = 58, p-value = 0.003054
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.5751638 -0.1351660
## sample estimates:
##      cor
## -0.3761795

with(cam.data,cor.test(drios,tree_1000m ))

##
## Pearson's product-moment correlation
##
## data:  drios and tree_1000m
## t = -1.7981, df = 58, p-value = 0.07737
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.45703763  0.02564122
## sample estimates:
##      cor
## -0.2297796
```

Influence of human populations / conucos

The Pemón are the only indigenous people inhabiting the Gran Sabana. There are four communities within the study area: Kawi (1100 m; - 61.243 W; 5.451 N; 50 people 2016), Mare-Paru (884 m; - 61.184 W; 5.594 N; 45 people in 2016), Uroy-Uaray (1,093 m; - 61.232 W; 5.442 N; 150 people in 2016) and Wuarapata (896 m; - 61.157; W 5.512 N; 50 people in 2016; information about the number of inhabitants was obtained from community leaders or capitanes).

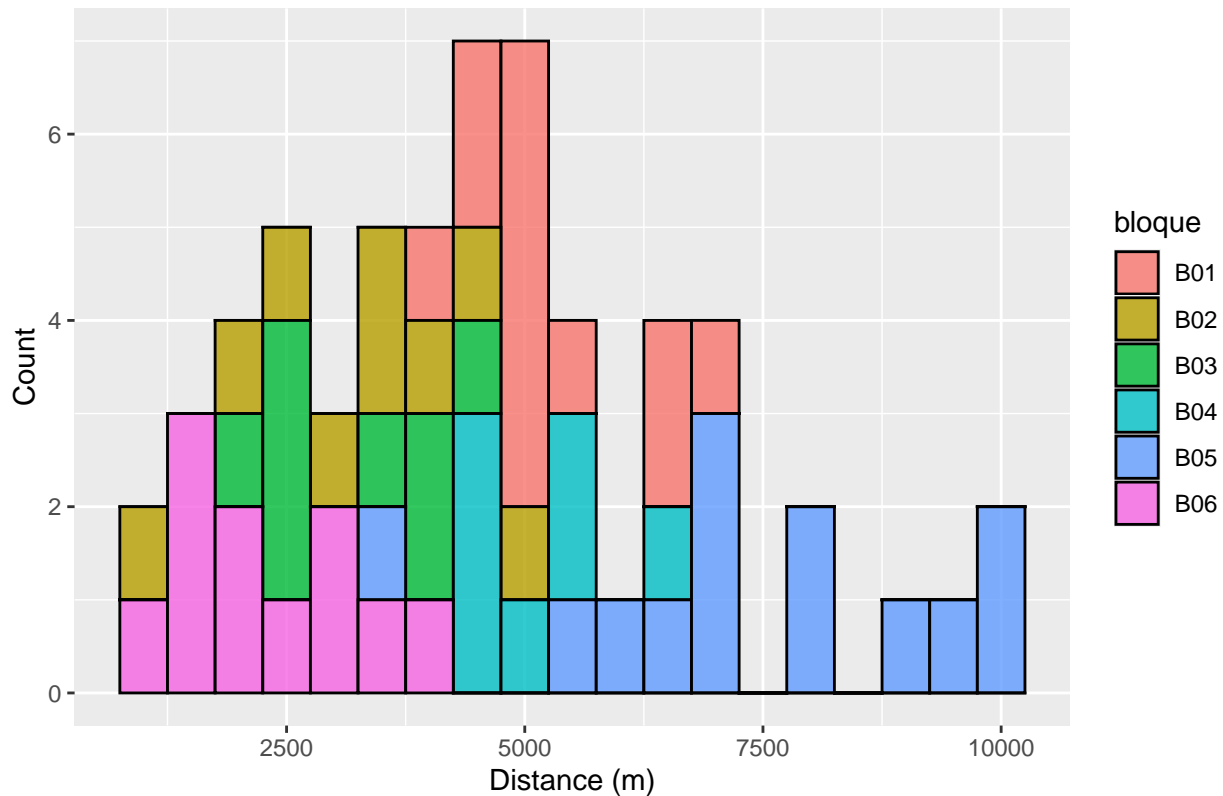
We calculate the distance from each camera to the nearest community, blocks 1 and 5 are furthest away from the communities:

```
d1 <- pointDistance(cam.data[,c("lon","lat")],
  coordinates(comunidades), lonlat=T, allpairs=T)

cam.data$dcom <- apply(d1,1,min)

ggplot(cam.data, aes(x=dcom,fill=bloque)) +
  geom_histogram(binwidth=500, alpha = .8,col='black') +
  labs(title="Histogram for distance to communities") +
  labs(x="Distance (m)", y="Count")
```

Histogram for distance to communities



During fieldwork we marked with a GPS the location of active and recently abandoned conucos (n=25) identified in situ and hunting sites (n=32) reported by interviewees and confirmed by the local guides.

```
table(coordinates((conucos))[,1]>-61.3) #
```

```
##
## FALSE TRUE
##    15    25
```

```
table(cam.data$ hunting)
```

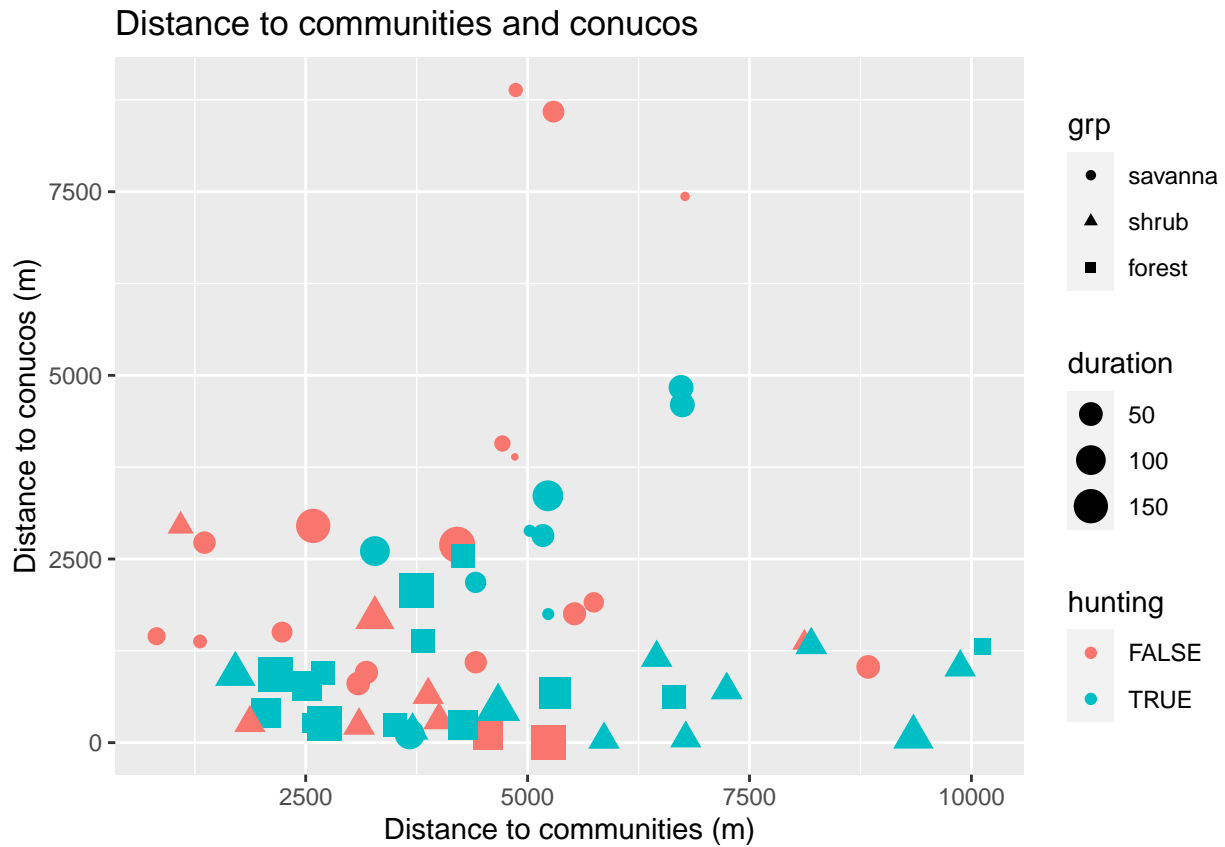
```
##
## FALSE TRUE
##    26    34
```

Distance from each camera to the nearest conuco was calculated using the GPS coordinates from cameras and conucos. This variable had an asymmetric distribution with a mean value of 1.58 km and a range from 0 to 8 km. We also recorded which cameras were located adjacent or near reported hunting sites (binomial variable hunting, FALSE n = 23, TRUE n = 34). Notice that three cameras are located in savanna habitat more than 5km from the nearest conuco.

```
d1 <- pointDistance(cam.data[,c("lon","lat")],
  coordinates(conucos)[,1:2], lonlat=T, allpairs=T)
```

```
cam.data$dcon <- apply(d1,1,min)
```

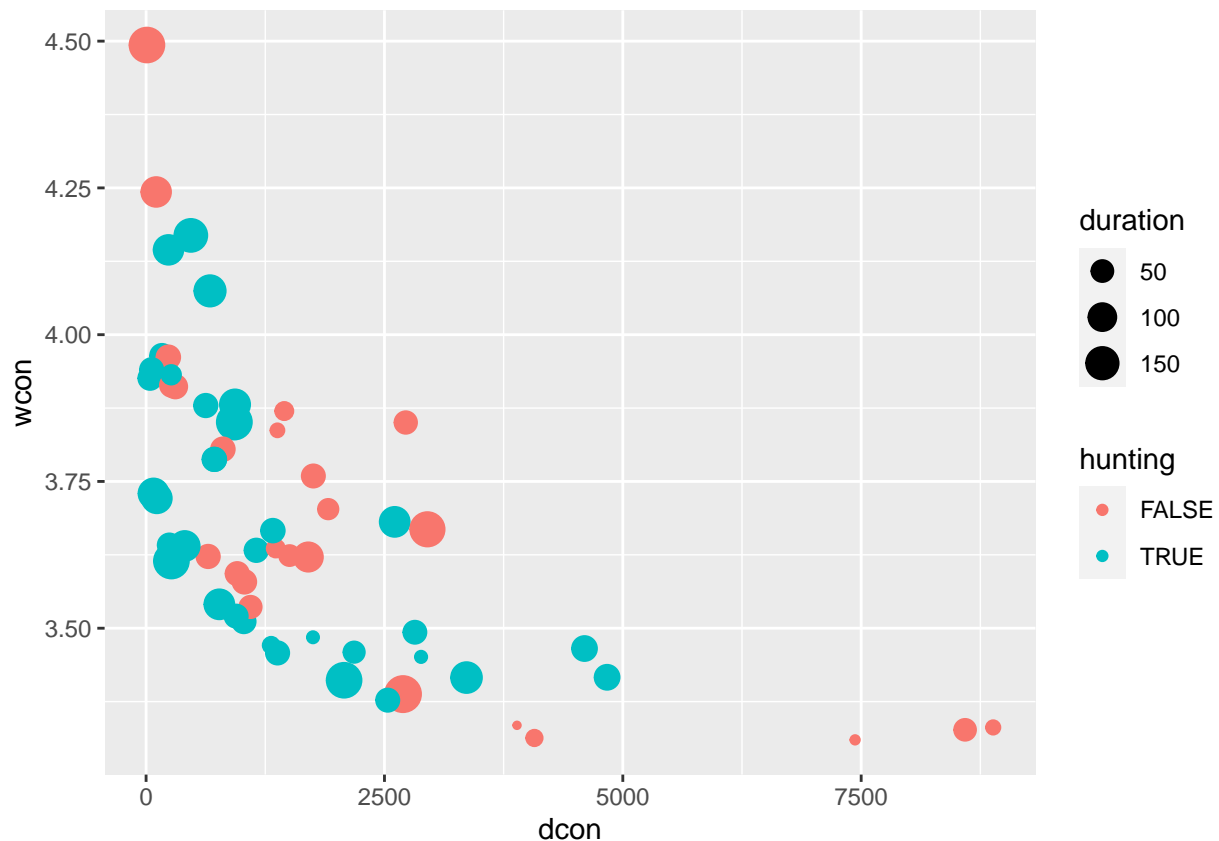
```
ggplot(cam.data,aes(x=dcom,y=dcon,shape=grp,colour=hunting,size=duration)) +
  geom_point() + labs(title="Distance to communities and conucos") +
  labs(y="Distance to conucos (m)", x="Distance to communities (m)")
```



As an alternative measure of the influence of conuco, we calculate density as $g(u) = (\sum w[i])$ where the weights are the inverse p-th powers of distance, $w[i] = 1/d(u, x[i])^p$ where $d(u, x[i])$ is the Euclidean distance from u to x[i].

```
p <- 0.25
w <- 1/((d1)^p)
cam.data$wcon <- apply(w,1,sum)

ggplot(cam.data,
  aes(y=wcon,x=dcon,colour=hunting,size=duration)) +
geom_point()
```

We check their correlation

```
cam.data %>% select(dcon,wcon,dcom) %>% cor()
```

```
##           dcon          wcon          dcom
## dcon  1.00000000 -0.6362453  0.08880916
## wcon -0.63624528  1.0000000 -0.17406593
## dcom  0.08880916 -0.1740659  1.00000000
```

```
# distance to conuco and density are neg. correlated
with(cam.data,cor.test(dcon,wcon))
```

```
##
## Pearson's product-moment correlation
##
## data: dcon and wcon
## t = -6.2807, df = 58, p-value = 4.674e-08
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.7663584 -0.4559895
## sample estimates:
## cor
## -0.6362453
```

```
# distance to conuco and dist. to communities are not sig. correlated
with(cam.data,cor.test(dcom,dcon))
```

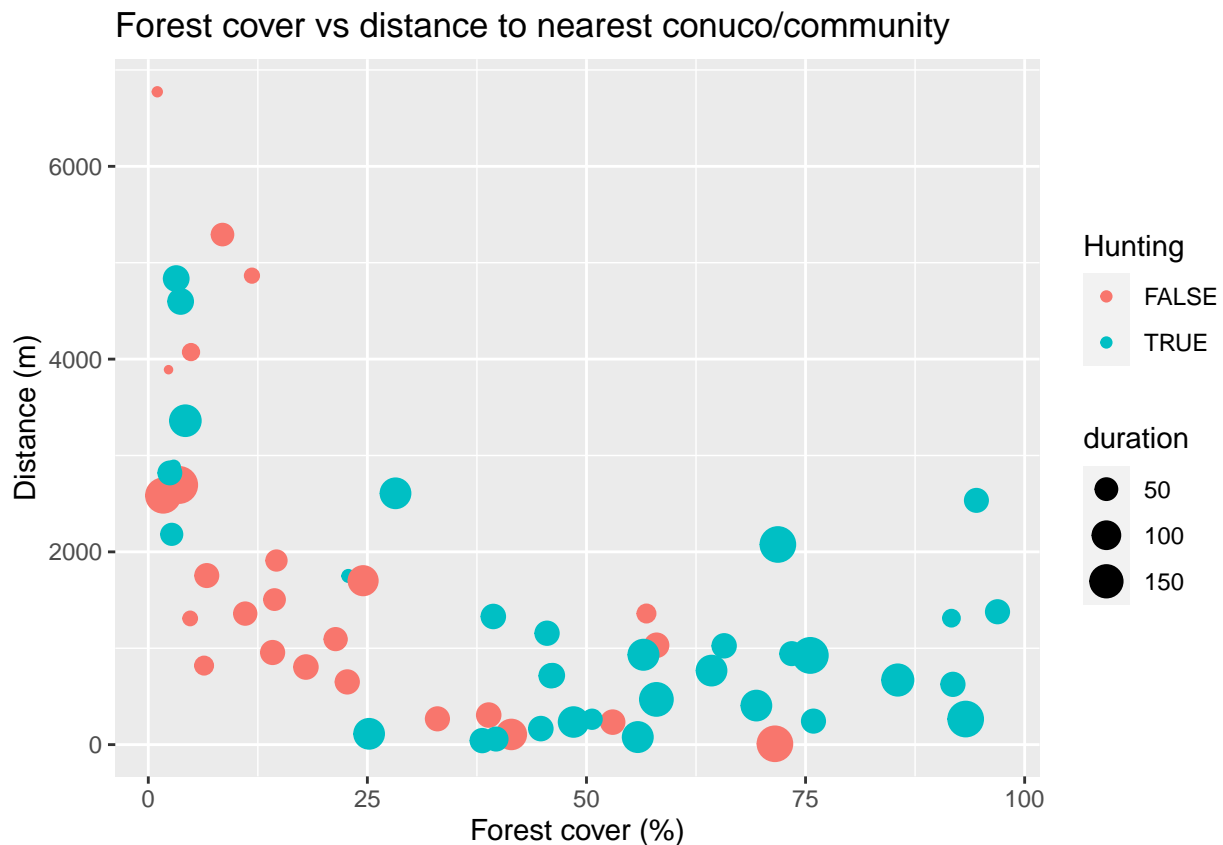
```
##
## Pearson's product-moment correlation
##
```

```
## data: dcom and dcon
## t = 0.67903, df = 58, p-value = 0.4998
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.1689249 0.3351754
## sample estimates:
##      cor
## 0.08880916
```

Here we plot distance to nearest conuco/community against tree cover:

```
cam.data$dhum <- with(cam.data, ifelse(dcon < dcom, dcon, dcom))

ggplot(data=cam.data, aes(y=dhum, x=tree_1000m, colour=hunting, size=duration)) +
  geom_point() +
  labs(title="Forest cover vs distance to nearest conuco/community") +
  labs(y="Distance (m)", x="Forest cover (%)", colour="Hunting")
```



```
##ggsave("Fig-bosque-caceria.pdf")
```

Signs of animal activity

We also recorded direct observations and indirect evidence (scats, tracks, scratches on trees, burrows, etc) of animal presence along the routes walked during field work, and recorded their coordinates with GPS. We had a total of $n=159$ records during 29 days of camera deployment and maintenance, with a mean of 16.8 km walked each day.

```
table(subset(eventos,bloque %in% sprintf("B%02i",1:6))$camara %in% "RAS")
```

```
##
## FALSE TRUE
## 920 159
```

```
field.walk <- subset(track_points,coordinates(track_points)[,1]>-61.3)
field.xy <- spTransform(field.walk,crs("+proj=utm +zone=19n"))
walk.dist <- pointDistance(coordinates(field.xy)[-length(field.xy),], coordinates(field.xy)[-1,], lonlat=FALSE)
walk.effort <- aggregate(walk.dist/1000,list(field.walk$time[-1]),sum)

summary(walk.effort$x)
```

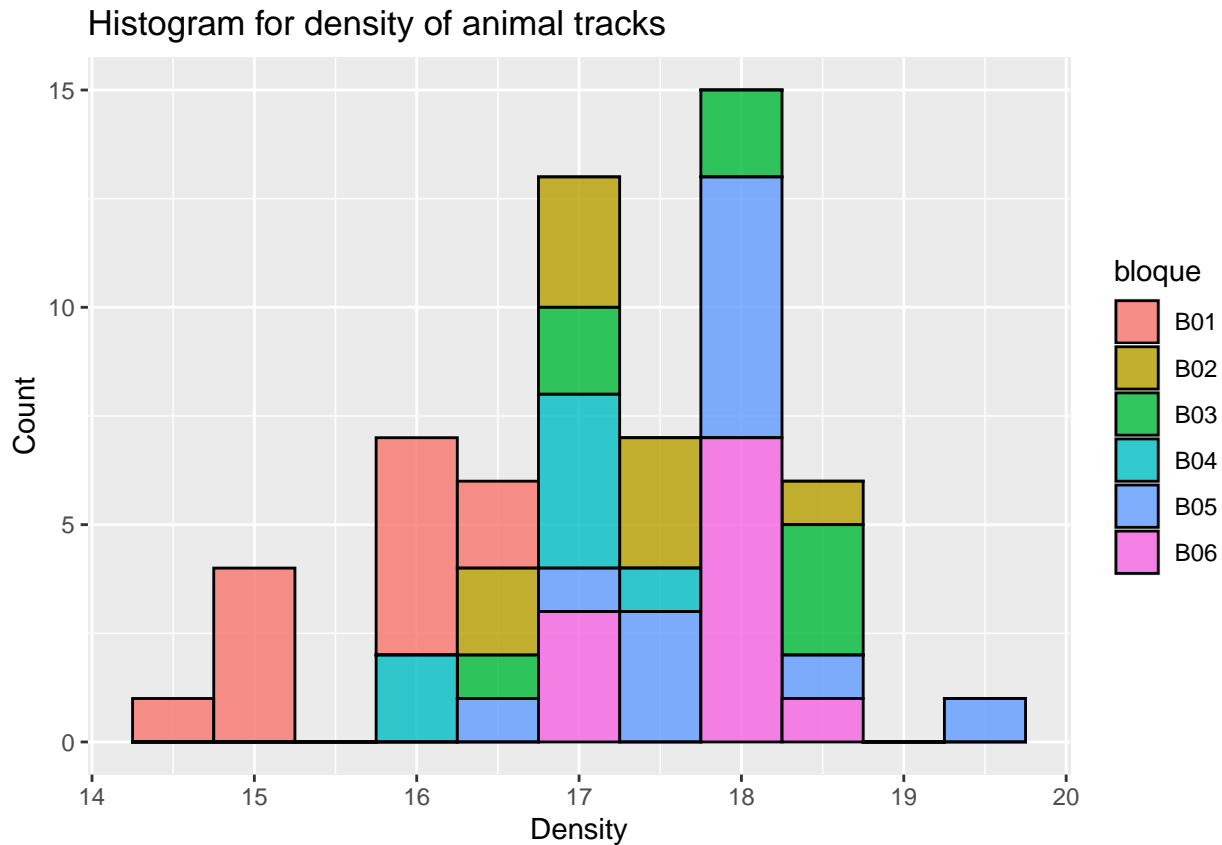
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.6851 11.6386 16.9653 16.8187 24.2906 34.3734
```

We would expect that detection will be higher near areas of more animal activity. We calculate density of animal observations and tracks as $g(u) = (\text{sum of } w[i])$ where the weights are the inverse p-th powers of distance, $w[i] = 1/d(u, x[i])^p$ where $d(u, x[i])$ is the Euclidean distance from u to $x[i]$.

```
draustos <- pointDistance(subset(eventos,camara %in% "RAS")[,c("long","lat")], cam.data[,c("lon","lat")])

p <- 0.25
w <- 1/((draustos)^p)
cam.data$dras <- apply(w,2,sum)

ggplot(cam.data,aes(x=dras,fill=bloque)) +
  geom_histogram(binwidth=.5, alpha = .8,col='black') +
  labs(title="Histogram for density of animal tracks") +
  labs(x="Density", y="Count")
```



Summarizing detections from the camera trap survey

Total sampling effort was 4,548 cameras per day, resulting in 7,466 captures, which translated into 771 detection events for mammals and 226 detection events for birds across 86 sampling units (Table 2 in Stachowicz et al., 2020). We identified mammal species (Eisenberg, 1989; Linares, 1998) and birds (Hilty, Tudor & Gwynne, 2003) using reference works for Venezuela and South America. We can summarize the number of events, fotos and individuals for each species:

```
eventos %>% mutate(on.camera=!camara %in% "RAS") %>%
  group_by(species,on.camera) %>%
  summarise(events=n(),fotos=sum(fotos),inds=sum(number.of.animals))
```

```
## `summarise()` regrouping output by 'species' (override with `.groups` argument)
```

```
## # A tibble: 46 x 5
```

```
## # Groups:   species [29]
```

##	species	on.camera	events	fotos	inds
##	<fct>	<lgl>	<int>	<int>	<int>
##	1 C.alector	TRUE	48	777	84
##	2 C.olivaceus	TRUE	8	42	8
##	3 C.paca	FALSE	7	0	7
##	4 C.paca	TRUE	265	1916	268
##	5 C.thous	FALSE	32	0	32
##	6 C.thous	TRUE	44	246	49
##	7 C.unicinctus	FALSE	2	0	2
##	8 C.unicinctus	TRUE	2	33	2
##	9 D.imperfecta	TRUE	14	51	14

```
## 10 D.kappleri FALSE 18 0 18
## # ... with 36 more rows
```

We now filter the detection events registered on cameras for a single species (here *Dasyprocta leporina*):

```
eventos$cdg <- as.character(camaras$ID.original)[
  match(paste(eventos$bloque,eventos$periodo,eventos$camara),
    paste(camaras$bloque,camaras$period,camaras$camera))]

eventos %>% mutate(f1 = chron(dates.=sprintf("%s-%s-%s",ano,mes,dia),
  times.=as.character(hora.ini),
  format = c(dates = "y-m-d", times = "h:m:s")),
  f2 = chron(dates.=sprintf("%s-%s-%s",ano,mes,dia),
  times.=as.character(hora.ini),
  format = c(dates = "y-mon-d", times = "h:m:s"),
  out.format = c(dates = "y-m-d", times = "h:m:s"))) %>%
transmute(cdg, camara, fotos, species=as.character(species), number.of.animals,
  fecha=chron(ifelse(is.na(f2),f1,f2),
  format = c(dates = "y-m-d", times = "h:m:s"))) %>%
filter(cdg %in% cam.data$cdg & species %in% "D.leporina") ->
  event.data

event.data %>%
  mutate(on.camera=!camara %in% "RAS") %>%
  group_by(species,on.camera) %>%
  summarise(events=n(),fotos=sum(fotos),inds=sum(number.of.animals))
```

```
## `summarise()` regrouping output by 'species' (override with `.groups` argument)
```

```
## # A tibble: 1 x 5
## # Groups:   species [1]
##   species    on.camera events fotos inds
##   <chr>      <lgl>      <int> <int> <int>
## 1 D.leporina TRUE      191  1423  192
```

To fit the model, we have to divide the continuous camera record (from camera activation to deactivation) into visits of fixed length. So we define a simple function to transform the list of events to a matrix:

```
make.obs.matrix <- function(x,y,w=NULL) {
  mtz <- matrix(0,nrow=length(unique(x$cdg)),
    ncol=length(y)-1,
    dimnames=list(unique(x$cdg),as.character(y)[-1]))
  for (k in 1:nrow(x)) {
    mtz[ x[k,"cdg"],] <-
      mtz[ x[k,"cdg"],] +
      table(cut(seq(x$fecha1[k],x$fecha2[k],by=1),
        breaks=y,label=as.character(y)[-1]))
  }
  if (!is.null(w)) {
    w %>% mutate(sessions=cut(fecha,breaks=y,label=as.character(y)[-1])) -> z
    mtz[mtz==0] <- NA
    mtz <- mtz*0
    for (k in seq(along=z$species)) {
      mtz[z[k,"cdg"],z[k,"sessions"]] <-
        mtz[z[k,"cdg"],z[k,"sessions"]] +
        z[k,"number.of.animals"]
    }
  }
}
```

```

}
  return(mtz)
}

```

We set a start date on the 21 september 2015, and select visit duration to create the matrix of sampling effort and observations. Initially we used a visit duration of one week, but this yields a large matrix with several zeros (no observations) and *NAs* (camera inactive):

```

ini <- chron(dates="2015-09-21",times="00:00:00",
  format = c(dates = "y-m-d", times = "h:m:s"))
visits <- ini + seq(from=7,by=7,length.out=28)

obs <- make.obs.matrix(data.frame(cam.data),visits,data.frame(event.data))
table(obs,useNA='always')

```

```

## obs
##    0     1     2     3     4     5     6     8     9    11 <NA>
## 598   62   25     7     2     1     3     1     1     1  919

```

We test larger visit duration to balance the number of zeros in the observation matrix, but this also reduces the number of detections. In fact we are “degrading” the input data since we are aggregating all detections per combination of site/visit, so multiple events get reduced to a single value of “1” (Kery and Royle 2016).

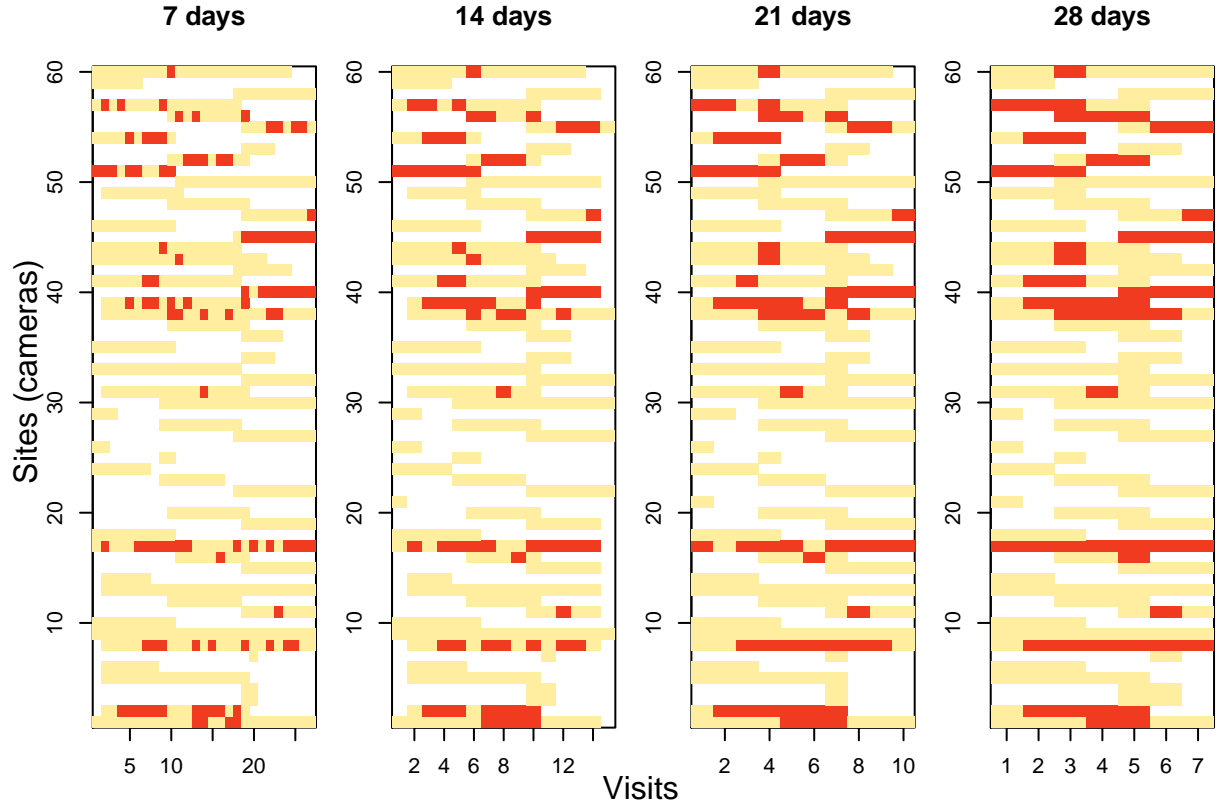
```

visits2 <- ini + seq(from=0,to=210,by=14)
visits3 <- ini + seq(from=0,to=210,by=21)
visits4 <- ini + seq(from=0,to=210,by=28)
clrs <- brewer.pal(3,'YlOrRd')

layout(matrix(1:4,ncol=4))
par(oma=c(2,2,0,0),mar=c(2,2,4,1))
obs <- make.obs.matrix(data.frame(cam.data),visits,data.frame(event.data))
image(1:ncol(obs),1:nrow(obs),t(obs>0),pty='m',col=clrs,xlab="",ylab="",main='7 days')
obs <- make.obs.matrix(data.frame(cam.data),visits2,data.frame(event.data))
image(1:ncol(obs),1:nrow(obs),t(obs>0),pty='m',col=clrs,xlab="",ylab="",main='14 days')
obs <- make.obs.matrix(data.frame(cam.data),visits3,data.frame(event.data))
image(1:ncol(obs),1:nrow(obs),t(obs>0),pty='m',col=clrs,xlab="",ylab="",main='21 days')
obs <- make.obs.matrix(data.frame(cam.data),visits4,data.frame(event.data))
image(1:ncol(obs),1:nrow(obs),t(obs>0),pty='m',col=clrs,xlab="",ylab="",main='28 days')

mtext('Visits',1,outer=T)
mtext('Sites (cameras)',2,outer=T)

```



Model fitting

For each species we fit a hierarchical Bernoulli/Poisson N-mixture model (Royle–Nichols models, RN-models: Royle & Nichols, 2003) to evaluate how the probability of occupancy relates to forest cover and distance to conuco, allowing for abundance-induced heterogeneity in detection probability. RN-models are based on the assumption that the detection probability at a survey point p_{ij} depends on the species’ site-specific abundance N_i :

$$p_{ij} = 1 - (1 - r_{ij})^{N_i}$$

where r_{ij} is the detection probability of a single individual. Repeated visits at a survey point generate a history of detection/nondetection events y_{ij} , from which p_{ij} is estimated. The abundance state (N_i) of site i was modeled as $N_i \text{ Poisson}(\lambda_i)$, while the observation process was modeled as $y_{ij}/N_i \text{ Bernoulli}(p_{ij})$. Estimation of p_{ij} allows us to draw conclusions about N_i .

In order to build detection histories for species recorded during the camera trap survey, we considered each camera location as a “site” (i ; 57 in total). We divided the total sampling period of 180 days into several “visits” (j). Here we use a visit duration of 21 days.

Formatting input

Now we arrange all data into a `unmarkedFrameOccu` object, we start with the detection history and observation covariates:

```
## Detection history
obs <- make.obs.matrix(data.frame(cam.data),visits3,data.frame(event.data))

## Observation covariates:
```

```
## sampling effort
sfrz <- make.obs.matrix(data.frame(cam.data),visits3)

## observation date
x <- seq(-1,1,length=ncol(obs))
obsDate <- matrix(rep(x,nrow(obs)),nrow=nrow(obs),byrow=T)
```

Now the site covariates:

```
sC <- data.frame(cam.data[match(rownames(obs),cam.data$cdg),
  c("bloque","H","h","dcon","dcom","wcon",
    "dhum","tree_0500m","tree_1000m","tree_2500m","tree_5000m","drios","ndvi","grp","dras","hunting"),
sC$bloque <- droplevels(sC$bloque)
## we select all camaras up to 5km from the conucos
ss <- sC$dcon<5000
```

We will standardize some variables:

```
for (k in c("H","h","dcon","dras","ndvi","tree_0500m","tree_1000m","tree_2500m","tree_5000m","drios","dhum"))
  sC[,k] <- (sC[,k]-mean(sC[,k]))/sd(sC[,k])
}
```

Now everything together:

```
UMF <- unmarkedFrameOccu((obs[ss,]>0)+0,
  siteCovs=sC[ss,,drop=F],
  obsCovs=list(date=obsDate[ss,],sfrz=sfrz[ss,]/21))
```

Explore alternative models

First we check all possible combinations of covariates for detectability for a fixed null model of abundance:

```
fm00 <- occuRN(~ bloque+H+h+date+dras+sfrz ~ H+h+grp, UMF,K=50)
if (!exists("all.combs.1"))
  all.combs.1 <- dredge(fm00,fixed=c("lam(grp)","lam(h)","lam(H)"))
sw(all.combs.1)
```

```
##               lam(grp) lam(h) lam(H) p(sfrz) p(H) p(h) p(date) p(dras)
## Sum of weights:   1.00    1.00  1.00  0.96  0.48 0.39 0.27  0.21
## N containing models: 64      64   64   32   32  32  32   32
##               p(bloque)
## Sum of weights:   0.04
## N containing models: 32
```

This test suggest that *effort* ($p(sfrz)$) is the detectability covariate with most support, and there is negligible support to covariate *blocks*, but the combination of *H* or *h* has moderate support. We check if this holds with a different covariate of abundance:

```
fm01 <- occuRN(~ bloque+H+h+date+dras+sfrz ~ tree_1000m, UMF,K=50)
if (!exists("all.combs.2"))
  all.combs.2 <- dredge(fm01,fixed=c("lam(tree_1000m)"))
sw(all.combs.2)
```

```
##               lam(ndvi) p(sfrz) p(H) p(date) p(dras) p(h) p(bloque)
## Sum of weights:   1.00    0.97  0.33 0.31  0.27  0.24 0.11
## N containing models: 64      32   32  32   32   32  32
```

... and found a similar result for *effort*, but less support for *H* or *h*.

The best ranking models include both *effort* and *H*, but there are other combinations with similar value of AIC.

```
subset(all.combs.1,delta < 2)

## Global model call: occuRN(formula = ~bloque + H + h + date + dras + sfrz ~ H + h +
##   grp, data = UMF, K = 50)
## ---
## Model selection table
##   p(Int) lam(Int) p(dat)    p(h)    p(H) p(sfr) lam(grp) lam(h)    lam(H) df
## 49 -2.111  -2.272                0.8789  1.703      + 0.3784 -0.60220  8
## 33 -2.393  -2.006                1.714      + 0.3331 -0.05924  7
## 41 -2.156  -2.204           -0.8305      1.591      + 0.9726  0.05041  8
## 57 -1.951  -2.322           -0.4304  0.8085  1.642      + 0.6511 -0.52690  9
## 51 -2.123  -2.251  0.2914           0.9031  1.752      + 0.3594 -0.61560  9
##   logLik  AICc delta weight
## 49 -105.803 230.6  0.00  0.296
## 33 -107.433 231.2  0.54  0.226
## 41 -106.131 231.3  0.66  0.214
## 57 -105.079 232.0  1.38  0.149
## 51 -105.332 232.5  1.89  0.115
## Models ranked by AICc(x)

subset(all.combs.2,delta < 2)
```

```
## Global model call: occuRN(formula = ~bloque + H + h + date + dras + sfrz ~ ndvi,
##   data = UMF, K = 50)
## ---
## Model selection table
##   p(Int) lam(Int) p(dat) p(drs)    p(H) p(sfr) lam(ndv) df   logLik  AICc delta
## 33 -2.423  -0.8525                1.690    1.465  4 -106.689 222.1  0.00
## 49 -2.259  -0.8726                0.3339  1.672    1.445  5 -106.032 223.2  1.09
## 35 -2.436  -0.8418  0.2608                1.742    1.434  5 -106.289 223.8  1.61
## 37 -2.490  -0.8513           0.2488                1.678    1.450  5 -106.426 224.0  1.88
##   weight
## 33  0.414
## 49  0.239
## 35  0.185
## 37  0.162
## Models ranked by AICc(x)
```

In order to compare the covariates of abundance, we take a slightly different approach, given the high correlation between variables related to habitat type and vegetation:

```
mdls <- list(
  "H+h" = occuRN(~ date+sfrz+H ~ H+h, UMF,K=50),
  "H+h+habitat" = occuRN(~ date+sfrz+H ~ H+h+grp, UMF,K=50),
  "H+h+treecover 500m" = occuRN(~ date+sfrz+H ~ H+h+tree_0500m, UMF,K=50),
  "H+h+treecover 1k" = occuRN(~ date+sfrz+H ~ H+h+tree_1000m, UMF,K=50),
  "H+h+treecover 2,5k" = occuRN(~ date+sfrz+H ~ H+h+tree_2500m, UMF,K=50),
  "H+h+treecover 5k" = occuRN(~ date+sfrz+H ~ H+h+tree_5000m, UMF,K=50),
  "H+h+ndvi" = occuRN(~ date+sfrz+H ~ H+h+ndvi, UMF,K=50),
  "habitat" = occuRN(~ date+sfrz+H ~ grp, UMF,K=50),
  "treecover 500m" = occuRN(~ date+sfrz+H ~ tree_0500m, UMF,K=50),
  "treecover 1k" = occuRN(~ date+sfrz+H ~ tree_1000m, UMF,K=50),
  "treecover 2.5k" = occuRN(~ date+sfrz+H ~ tree_2500m, UMF,K=50),
```

```

"treecover 5k" = occuRN(~ date+sfrz+H ~ tree_5000m, UMF,K=50),
"ndvi" = occuRN(~ date+sfrz+H ~ ndvi, UMF,K=50)
)

aictab(mdls,second.ord=T)

```

```

##
## Model selection based on AICc:
##
##           K   AICc Delta_AICc AICcWt Cum.Wt      LL
## ndvi          6 224.88      0.00   0.38   0.38 -105.60
## treecover 1k   6 225.15      0.28   0.33   0.71 -105.74
## treecover 2.5k 6 227.61      2.73   0.10   0.81 -106.96
## treecover 500m 6 228.65      3.77   0.06   0.86 -107.49
## H+h+ndvi       8 228.92      4.04   0.05   0.91 -104.96
## H+h+treecover 1k 8 229.77      4.89   0.03   0.95 -105.38
## habitat        7 230.98      6.10   0.02   0.96 -107.35
## H+h+treecover 2,5k 8 232.20      7.32   0.01   0.97 -106.60
## H+h+treecover 500m 8 232.32      7.44   0.01   0.98 -106.66
## H+h+habitat     9 232.49      7.62   0.01   0.99 -105.33
## treecover 5k    6 233.35      8.47   0.01   1.00 -109.83
## H+h            7 236.32     11.44   0.00   1.00 -110.02
## H+h+treecover 5k 8 236.98     12.10   0.00   1.00 -108.99

```

Model with NDVI or tree cover at 1km are slightly better than other alternatives.

Now rivers:

```

mdls <- list(
  "treecover 500m" = occuRN(~ date+sfrz+H ~ tree_0500m, UMF,K=50),
  "treecover 1k" = occuRN(~ date+sfrz+H ~ tree_1000m, UMF,K=50),
  "treecover 2.5k" = occuRN(~ date+sfrz+H ~ tree_2500m, UMF,K=50),
  "treecover 5k" = occuRN(~ date+sfrz+H ~ tree_5000m, UMF,K=50),
  "treecover 500m + rios" = occuRN(~ date+sfrz+H ~ tree_0500m+drios, UMF,K=50),
  "treecover 1k + rios" = occuRN(~ date+sfrz+H ~ tree_1000m+drios, UMF,K=50),
  "treecover 2.5k + rios" = occuRN(~ date+sfrz+H ~ tree_2500m+drios, UMF,K=50),
  "treecover 5k + rios" = occuRN(~ date+sfrz+H ~ tree_5000m+drios, UMF,K=50)
)

aictab(mdls,second.ord=T)

```

```

##
## Model selection based on AICc:
##
##           K   AICc Delta_AICc AICcWt Cum.Wt      LL
## treecover 1k + rios 7 224.25      0.00   0.41   0.41 -103.98
## treecover 1k       6 225.15      0.91   0.26   0.68 -105.74
## treecover 2.5k + rios 7 227.13      2.88   0.10   0.78 -105.42
## treecover 500m + rios 7 227.26      3.01   0.09   0.87 -105.49
## treecover 2.5k     6 227.61      3.36   0.08   0.94 -106.96
## treecover 500m     6 228.65      4.40   0.05   0.99 -107.49
## treecover 5k + rios 7 232.98      8.73   0.01   1.00 -108.35
## treecover 5k       6 233.35      9.10   0.00   1.00 -109.83

```

Slightly better with rios...

Models for all different measures of human influence (conuco density, distance, distance to communities and combined conuco/community distance) provide very similar results when using one variable at a time.

```
fm11 <- occuRN(~ date+sfrz+H ~ tree_1000m+dcom, UMF,K=50)
fm12 <- occuRN(~ date+sfrz+H ~ tree_1000m+dcon, UMF,K=50)
fm13 <- occuRN(~ date+sfrz+H ~ tree_1000m+wcon, UMF,K=50)
fm14 <- occuRN(~ date+sfrz+H ~ tree_1000m+dhum, UMF,K=50)
fm15 <- occuRN(~ date+sfrz+H ~ tree_1000m+dcon+dcom, UMF,K=50)
fm16 <- occuRN(~ date+sfrz+H ~ tree_1000m+dcon+wcon, UMF,K=50)
fm17 <- occuRN(~ date+sfrz+H ~ tree_1000m+dcon*wcon, UMF,K=50)

aictab(list(fm11,fm12,fm13,fm14,fm15,fm16,fm17),
modnames=c("comunity","conuco (dist)","conuco (dens)","comunity/conuco combined","comunity + conuco","conuco (dist*dens)"))
```

```
##
## Model selection based on AICc:
##
##           K   AICc Delta_AICc AICcWt Cum.Wt      LL
## conuco (dist)      7 225.37      0.00  0.31  0.31 -104.54
## comunity/conuco combined 7 225.71      0.34  0.26  0.57 -104.71
## conuco (dist+dens)  8 227.37      2.00  0.11  0.68 -104.19
## comunity           7 227.47      2.09  0.11  0.79 -105.59
## conuco (dens)      7 227.68      2.30  0.10  0.88 -105.70
## comunity + conuco  8 228.01      2.64  0.08  0.96 -104.51
## conuco (dist*dens)  9 229.67      4.30  0.04  1.00 -103.92
```

Distance to conuco is slightly better than other measures, but there is a strong correlation between them:

```
cor.test(cam.data$dcon,cam.data$tree_1000m)
```

```
##
## Pearson's product-moment correlation
##
## data: cam.data$dcon and cam.data$tree_1000m
## t = -4.645, df = 58, p-value = 1.999e-05
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.6841717 -0.3074330
## sample estimates:
## cor
## -0.5207087
```

While distance to community is not:

```
cor.test(cam.data$dcom,cam.data$tree_1000m)
```

```
##
## Pearson's product-moment correlation
##
## data: cam.data$dcom and cam.data$tree_1000m
## t = 1.0646, df = 58, p-value = 0.2915
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.1196910 0.3790413
## sample estimates:
## cor
## 0.1384413
```

For this study we calculated tree cover and distance to nearest conuco as covariates of species abundance.

Covariates of N_i (site covariates) and p_{ij} (observation covariates) were modeled using the logit link.

To test our hypothesis we need a measure of influence of conucos while controlling for the influence of habitat on species abundance, and the spatial and temporal heterogeneity in probability of detection. We explored several covariates with alternative parameterizations to ensure best possible model fit given the restricted sample size and low number of detection for some species, (see Suppl. Material for details). As site covariates we used tree cover percentage around the camera trap (tree_buffer) and distance to nearest conuco (dist_conuco), both variables were standardized to zero mean and unit standard deviation. We used 1 km radius under the assumption that it is wide enough to represent the area of the most abundant game species home range (*Cuniculus paca* 2 - 3 ha, Jorge and Peres 2005; *Dasyprocta leporina* 3.4 - 1.6 ha, Benavides et al. 2017) and narrow enough to maintain variability in tree cover within the scale of a camera trapping site (Scotsont et al 2017). Since a couple of species might be associated with shrub habitat with intermediate values of tree cover (Stachowicz et al. 2020) we added a quadratic term (tree_buffer2) to their models.

We used sampling date, sampling effort (camera/day), and density of direct and indirect off-camera records to account for spatial and temporal heterogeneity in detectability (Cubaynes et al., 2010, REFs). Sampling date (date) was recorded as the number of days since the start of sampling (21 September 2015) to the beginning of the “visit” and standardized to zero mean and unit standard deviation. Sampling effort (effort) was calculated as the number of days the camera remained active divided by the duration of the visit. Thus, effort was always ≤ 1 , and was set to empty value (NA) when the camera was not present or inoperative during the whole duration of the visit. We calculated tracks density (tracks_dens) as the inverse distance weighted sum of wildlife activity recorded off-camera (direct observations and indirect evidence of animal presence during field-work): Eq. 4 $\text{track_dens}_i = \sum_j \frac{1}{d_{ij}^p}$, where d_{ij} is the distance between camera i and record j for all $k = 159$ records, and $p = 0.25$. The result was standardized to zero mean and unit standard deviation.

For each species, we fitted a full model including the three observation covariates (p (date + effort + tracks_dens)) and the two site covariates (λ (tree_buffer + tree_buffer2 + dist_conuco)) using the occuRN function of the R package unmarked (Fiske & Chandler, 2011). We assessed model fit for the full model using goodness of fit test based on Pearson χ^2 and parameter bootstrapping with 10,000 samples, and inspecting under- or overdispersion (c , calculated by dividing the observed Chisq statistic by the mean of the statistics obtained from bootstrap samples), magnitude of parameter estimates and standard errors, and predicted values of the state variable at the sample locations (MacKenzie and Bailey 2004; Royle and Nichols 2016). For species with a suitable full model, we proceeded to create a model selection table with all combinations of covariates (32 models for species with linear effect of tree cover and 48 for species with quadratic effect of tree cover), ranked models according to information criteria corrected for small sample size (AICc if $c \leq 1$ or QAICc if $c > 1$), and the corresponding $\delta(Q)AICc$ and model weights.

We assessed the relative importance of each detection and occupancy covariate by calculating the sum of weights of the model containing that variable (Burnham and Anderson 2002; Symonds and Moussalli 2011). Values range from zero to one indicating increasing levels of support, and we use an informal scale to describe the level of support as very strong (>0.9), strong (0.6 - 0.9), moderate (0.3 - 0.6) and low (<0.3). We further calculated model averaged coefficients and predictions of the state variable (λ) based on the subset of models with $\delta(Q)AICc \leq 10$ (Burnham et al 2011; Mazerolle 2020). In case of overdispersion ($c > 1$) we assumed the lack of fit is due to unaccounted sources of error and used the value of c to inflate the standard errors and confidence intervals. For underdispersed models ($c < 1$), no modification to standard errors or intervals was made, but consider these as conservative assessments of uncertainty (Kery and Royle 2019).

Selection of hunting localities

To assess hunting behaviour of the Pemon we evaluated both the interview responses on vegetation type and season with contingency tables, and the spatial information on reported hunting sites with a logistic regression. We tabulated the number of interview responses from each community for the three levels of preferred hunting vegetation types (forest, savanna and mixed) and the two levels of hunting seasons (dry, and rainy season). We used the (Chi-square) test to assess the significance of the relationship between variables. For the logistic

regression we used the data collected during field work at 57 sites with cameras and fitted a logistic regression to the binomial hunting variable with formula: $\text{Eq 5 } \text{logit}(\text{hunting}) = \beta_0 + \beta_1 \text{treebuffer} + \beta_2 \text{distance}$

notes on the occu RN model

Kery and Royle 2016:

The RN model may be useful to estimate abundance from replicated measurements of presence/absence, to accommodate detection heterogeneity when focusing on occupancy (Dorazio, 2007) or to link occupancy data with abundance data in an integrated model (Conroy et al., 2008)

In practice, the RN model has not been extensively used to actually estimate abundance. Arguably, it is more sensitive to parametric assumptions and it will fail when abundance and/or detection are high. Moreover, adopting mixture distributions other than the Poisson, such as the NB, leads to unstable estimates and often also to biologically unreasonable high estimates for abundance. Hence, the RN model may be most useful as an occupancy model that accounts for part of the site-specific heterogeneity in detection (Dorazio, 2007) or in integrated models to link occupancy data with data that are directly informative about abundance (Conroy et al., 2008). Note also that Yamaura et al. (2011) have used it as the basis for their community models (Chapter 11).