**Jamie Burke**

HW4.R

Step 1: Study organism: shorebirds

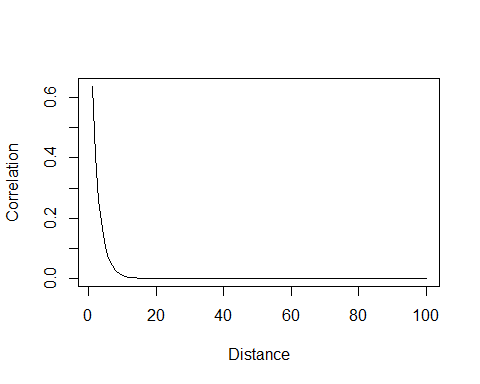
Step 2: Environmental predictor: % cover mudflat

library("raster")

library("rgeos")

library("rgdal")

# Define function to draw random samples from a multivariate normal  
# distribution  
  
rmvn <- function(n, mu = 0, V = matrix(1)) {  
 p <- length(mu)  
 if (any(is.na(match(dim(V), p))))   
 stop("Dimension problem!")  
 D <- chol(V)  
 t(matrix(rnorm(n \* p), ncol = p) %\*% D + rep(mu, rep(n, p)))  
}  
  
# Set up a square lattice region  
simgrid <- expand.grid(1:50, 1:50)  
n <- nrow(simgrid)  
  
  
# Set up distance matrix  
distance <- as.matrix(dist(simgrid))  
  
# Generate random variable  
  
phi = 0.45 #phi determines scale of distance variation  
#how does changing phi change the spatial aggregation in the plotted raster?  
plot(1:100, exp(-phi \* 1:100), type = "l", xlab = "Distance", ylab = "Correlation")



#X <-rmvn(1, rep(0, n),exp(-phi \* distance))  
  
#X is predictor variable. Does it make more sense to have a discrete or  
#continuous predictor?  
  
#X <- rpois(n, lambda=exp(-1+rmvn(1, rep(0, n), exp(-phi \* distance))))  
  
#or rbinom  
#X <- rbinom(n, prob=plogis(-1+rmvn(1, rep(0, n), exp(-phi \* distance))),size=3)  
#Xraster <- rasterFromXYZ(cbind(simgrid[, 1:2] - 0.5, X))  
  
#or beta  
  
mudflat=rmvn(1, rep(0, n), exp(-phi \* distance))  
  
#mean\_mudflat=plogis(-1+rmvn(1, rep(0, n), exp(-phi \* distance)))  
  
X<-rbeta(n,(shape1=plogis(mudflat)\*phi),shape2=plogis((1-mudflat))\*phi)  
  
  
# Visualize results  
Xraster<-rasterFromXYZ(cbind(simgrid[, 1:2] - 0.5, X))  
plot(Xraster)  
  
#Converting raster to a dataframe  
spat\_dat=rasterToPoints(Xraster)  
  
  
#how many points can you sample?  
GO=sample(x=c(1:nrow(spat\_dat)),size=200)  
#head(GO)  
  
points(spat\_dat[GO,c(1:2)],pch=19)

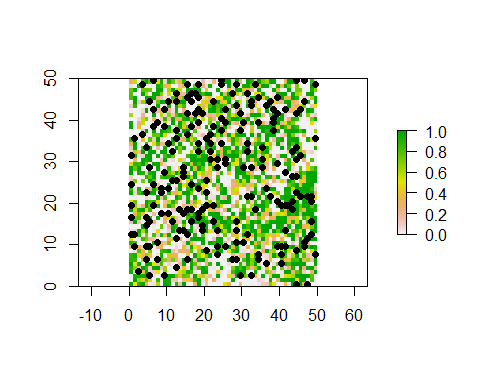
Step 3 output:

a. Area of mudflat is likely to predict counts of shorebirds because shorebirds spend most of their time feeding and mudflats are their preferred foraging environment.

b. My phi value is 0.45 because it is more likely that you will find mudflat areas closer to mudflat areas.

c. The number of points to sample is constrained to time in the field.

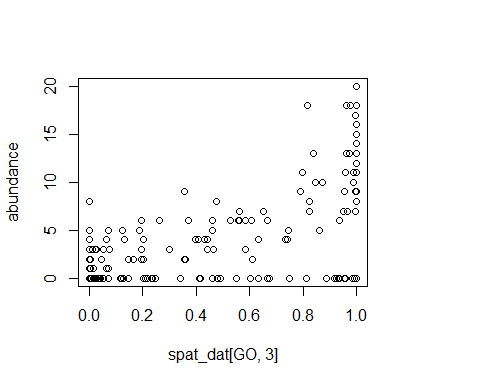
d.



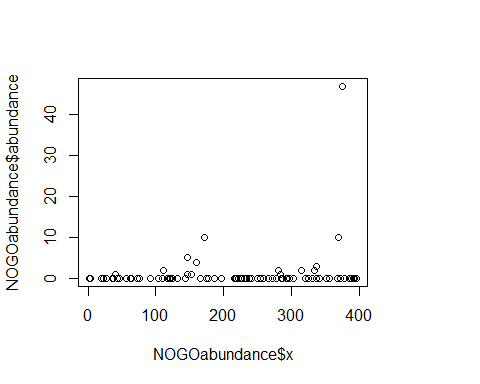
e. A reason you might observe a hurdle model is you have many zero counts and they may be for many different reasons and possibility other variables something like many shorebirds have already migrated, and some shorebirds tend to flock and stick together, and so random sampling might miss these groups.

f. By multiplying the binomial and the negative binomial variable to get a total abundance, we are assuming that when there is a count of zero in the presence/absence model that the organism is truly not there even though it may have been there but we didn’t count it.

#how would you deal with spatial autocorrelation here?  
  
presence\_intercept=0.1  
presence\_slope=1  
  
#assumptions?  
PA=rbinom(200,plogis(presence\_intercept+spat\_dat[GO,3]\*presence\_slope),  
 size=1)  
  
  
count\_intercept=0.5  
count\_slope=2  
over\_dispersion=20  
  
abundance=PA\*rnbinom(200,mu=  
 exp(count\_intercept+count\_slope\*spat\_dat[GO,3]),  
 size=over\_dispersion)  
  
plot(abundance~spat\_dat[GO,3])



#head(spat\_dat[GO,3])  
#head(abundance)  
  
#write.csv(data.frame(abundance,spat\_dat[GO,3]),file="HW4\_shorebirdfakedata\_Jamie.csv")  
  
####getting data######  
  
#PA1=ifelse(abundance>0,1,0)  
#glm(PA1~environmental\_covariate,family="binomial")  
#abundance2<-abundance[which(abundance>0),]  
#glm.nb(abundance2~environmental\_covariate2)  
  
  
NOGOabundance<-read.csv("NOGO\_abundance2.csv")  
  
plot(NOGOabundance$abundance~NOGOabundance$x)



####presence/absence######  
PA1=ifelse(NOGOabundance$abundance>0,1,0)  
NOGOmod<-glm(PA1~NOGOabundance$x, family="binomial")  
coef(NOGOmod)

## (Intercept) NOGOabundance$x   
## -1.902961319 0.001604192

plogis(-1.903)

## [1] 0.1297693

####abundance#####  
  
abundance2<-abundance[which(NOGOabundance$abundance>0)]  
environfactor<-NOGOabundance$x[which(NOGOabundance$abundance>0)]  
abun2mod<-glm(abundance2~environfactor, family="poisson")  
coef(abun2mod)

## (Intercept) environfactor   
## 1.27110706 -0.00055472

exp(3.5044)

## [1] 33.26148

library(MASS)

## Warning: package 'MASS' was built under R version 3.5.2

##   
## Attaching package: 'MASS'

## The following objects are masked from 'package:raster':  
##   
## area, select

abun2mod2<-glm.nb(abundance2~environfactor)  
coef(abun2mod2)

## (Intercept) environfactor   
## 1.298907937 -0.000677129

h. True PA intercept: -2.5, slope=1.5, True count intercept: 0.5, slope:0.75

The true intercept and slope for both the PA and abundance models were not captured.

i. You could use a plot to look at spatial autocorrelation of the response variable and use it to determine phi given the multivariate normal.