**Lab Assignment 5 – Fitting a Resource (Habitat) Selection Function**

**WFSC 570 Wildlife Habitat Analysis**

**Due October 29, 2024**

For this assignment, you will fit a resource (habitat) selection function (RSF) describing third-order habitat selection for our favorite fisher, Lupe. This lab will use the same data from Fieberg et al. (2021) that we have been using in class but with a few additional covariates. The goal of this lab is to have you fit a third-order multiscale RSF for Lupe and make inferences about scales of effect and habitat selection.

Your data for this lab include a single RDS file (*All\_Lupe\_data\_multiscale\_df.rds*) which includes environmental data at all 3004 used points and 300,400 available points:

* *y*
  + A binary covariate denoting used (0) and available (1) points
* *w*
  + The weights for each used (1) and available (5000) point to better allow our binomial GLM’s slopes to approximate the slopes of an Inhomogeneous Poisson Point Process (IPP) model.
* *X* and *Y*
  + The UTM coordinates for each point (not needed for this lab)
* *pt\_elev*
  + Elevation at each used and available point (this is the same as the Elevation covariate we have been using)
* *ForEdge\_100* through *ForEdge\_2000*
  + Proportion of circular buffers (100 – 2000 m radii) that is forest edge.
* *Grass\_100* through *Grass\_2000*
  + Proportion of circular buffers (100 – 2000 m radii) that is grassland land cover
* *Wet\_100* through *Wet\_2000*
  + Proportion of circular buffers (100 – 2000 m radii) that is wetland land cover
* *Elev\_100* through *Elev\_2000*
  + Mean elevation within circular buffers (100 – 2000 m radii)
* *TPI\_100* through *TPI\_2000*
  + Topographic position index calculated using 100 – 2000 m radius circular buffers
* *PopDen*
  + Human population density
* *Forest, Grassland, and Wetland*
  + Binary covariates denoting whether (1) or not (0) a point falls in forest, grassland, and wetland land covers.

*Note: You will need to load the R packages MuMIn, ggplot2, and car for this lab*

**Complete the following questions/tasks:**

**Task I: Perform a pseudo-optimization procedure to select the scale of effect for 1) Forest Edge, 2) Elevation, 3) TPI, 4) Mean Proportion of Grassland, and 5) Mean Proportion of Wetland**

Use the same procedure from the scale lab to identify the scales of effect for these five covariates. You can use the following code to simplify the procedure and create a plot showing how AIC changes across scales. We will add a slight twist here and instead of plot AIC as a function of scale we will plot Delta AIC as a function of scale. The Delta AIC is the difference between the lowest AIC in the candidate model set and the AIC of each model. The model with the best empirical support will have a Delta AIC = 0. Delta AIC allows us to say something about variation in support amongst models. For example, all models within some threshold level of Delta AIC (e.g., 2 or 4) are said to have similar degrees of empirical support for being the “best” model within a candidate model set.

covs <- c("ForEdge","Elev","TPI","Grass","Wet")

buffers <- seq(100,2000,by=100)

col\_names <- c("Covariate","Scale","AIC","Delta\_AIC","AICw")

scales <- as.data.frame(matrix(NA, nrow = length(covs)\*length(buffers),

ncol = length(col\_names)))

colnames(scales) <- col\_names

i.row <- 1

for(i in 1:length(covs)){

cov\_i <- covs[i]

cat("Starting",cov\_i,"\n")

data\_i <- test[,c("y","w",paste0(cov\_i,"\_",buffers))]

cov\_cols <- paste0(cov\_i,"\_",buffers)

for(j in cov\_cols){

tmp\_col <- j

tmp\_data <- data\_i[,c("y","w",tmp\_col)]

tmp\_model <- glm(tmp\_data[,1]~tmp\_data[,3],weights=tmp\_data[,2],family=binomial)

scales$Covariate[i.row] <- cov\_i

scales$AIC[i.row] <- AIC(tmp\_model)

scales$Scale[i.row] <- as.numeric(strsplit(j,"\_")[[1]][2])

cat("Finished scale",as.numeric(strsplit(j,"\_")[[1]][2]),"\n")

i.row <- i.row + 1

}

min\_i <- min(scales$AIC[which(scales$Covariate==cov\_i)])

delta\_i <- scales$AIC[which(scales$Covariate==cov\_i)] - min\_i

w\_i <- Weights(scales$AIC[which(scales$Covariate==cov\_i)])

scales$AICw[which(scales$Covariate==cov\_i)] <- w\_i

scales$Delta\_AIC[which(scales$Covariate==cov\_i)] <- delta\_i

}

scale\_plot <- ggplot(data=scales,

aes(x=Scale,y=Delta\_AIC))+

geom\_hline(yintercept = 0) +

geom\_line(color="black")+

geom\_point(color="black")+

facet\_wrap(~Covariate)+

theme\_bw()

scale\_plot

**Answer the following questions:**

1. **What is the scale of effect for each covariate?**
2. **What is the most common scale of effect across these covariates?**
3. **Describe in general terms the strength of the empirical support for each scale of effect. Do you see many scales having similar levels of support or does one scale predominately have all the support?**

**Task II: Fit a multi-covariate pseudo-optimized multi-scale model**

Once you have identified the scale of effect for each covariate, take those covariates at their scales of effect and combine them into a multi-covariate pseudo-optimized multi-scale model. Your model should include the following covariates:

* Elevation
* Forest Edge
* Proportion of Grassland
* Proportion of Wetland
* TPI
* Population Density

Population density will be included in the model at its original scale since the pixel size of the original raster was relatively large (so was the land cover raster but we will ignore that here). Before you include elevation, proportion of grassland, and proportion of wetland at their scales of effect, compare the AIC of each scale of effect to the AIC of a model fit using the covariate measured as the pixel value of the individual point. For example, does the elevation of the point itself (which we used in class) have a lower AIC than mean elevation within some circular buffer around the point? Use whichever has the lower AIC (the value at the point or within a buffer) in the final model.

Once you fit the model we should test for collinearity. Collinearity is something to generally avoid in regression models as it can inflate your standard errors and potentially bias your slope estimates. One nice way to test for collinearity is to use *variance inflation factors (VIF)*. Each covariate in the model has a VIF value and covariates with VIF’s > 3 are often discarded from the final model. Load the car library which has a function (*vif*) that you can use. You simply run *vif()* on your fitted binomial GLM object and the function will return the VIF for each covariate.

**Answer the following questions:**

1. **Did the value of elevation, grassland, and wetland have more empirical support at the point scale or at the scale of effect?**
2. **Did any covariates have VIF > 3? If two covariates had VIF > 3, fit two binomial GLMs, one with each of the two covariates, and see which has the greatest empirical support. Retain the covariate with the greatest empirical support in your final model and discard the other covariate.**
3. **Do you have any categorical covariates in your final model?**

**Task III: Make inferences from your final model**

**Answer the following questions:**

1. **What covariates were significant (α = 0.05) in the final model?**
2. **What covariates had the strongest effect sizes?**
3. **Select three covariates from your final model. For each covariate select two different values of that covariate. Holding all other values constant at their means, calculate how much more or less likely Lupe would be to select Location 2 relative to Location 1 for each of your three covariates.**

Your assignment is to work in pairs, complete the following tasks and answer the following questions as a pair, and submit a single written report for each pair describing how you completed the tasks and your answers to the questions. Reports should be written using complete sentences and paragraph structure. Also include your R script (either as a separate file and copied-and-pasted into the end of your report).

**This assignment will be due on D2L by the beginning of lab (2:00) on Tuesday October 29, 2024.**