**Lab Assignment 3 – Occupancy Models for Wildlife Habitat Analyses**

**WFSC 570 Wildlife Habitat Analysis**

**Due October 8, 2024**

For this assignment, you will analyze some wildlife-habitat data using the single-species single-season occupancy model of Mackenzie et al. (2002). You will fit these models using the *occu* function within the unmarked package in R and make some inferences about wildlife-habitat relationships.

These data are inspired by bobcat in Illinois. Bobcat declined in Illinois, and many other Midwestern states, due to overharvest and landscape changes (i.e., loss of native forest and prairie communities to row-crop agriculture). Bobcat were state listed as threatened in Illinois in 1977. In the following decades bobcat began expanding their distribution northward from southern Illinois such that by 2018 bobcat were fairly widespread in the western and southern part of Illinois (Bauder et al. 2023). Within the Midwest USA, bobcat are generally positively associated with grassland and forest land covers.

For this lab, we have data inspired by a bobcat camera trapping project in Illinois. 102 cameras were placed at sites across Illinois and deployed for five weeks resulting in five “surveys” or five “visits” per site. One survey-level covariate was collected: DOY (day-of-year) where 1 = 1 October and 45 = 14 November. Two site-level (camera-level) covariates were collected: Forest and Grassland which are the percentages of a 500-m radius buffer around each camera that was the respective land cover class.

The data for this lab assignment include three CSV files.

* *Bobcat\_encounter\_history.csv*
  + This data frame contains the detection/non-detection data, also known as the encounter history. Each row represents a single site (camera) and each column represents a different 7-day sampling period (i.e., survey or visit).
* *Bobcat\_site\_covariates.csv*
  + This data frame contains two columns, one representing the percent of the 500-m radius buffer that is forest land cover (Forest) and the other representing the percent of the 500-m radius buffer that is grassland land cover (Grassland).
* *Bobcat\_survey\_level\_covariates.csv*
  + This data frame denotes the day-of-year (DOY) of each 7-day sampling period. Notice that it has the same number of rows and columns as the encounter history.

*Note: This lab uses three R packages: unmarked, AICcmodavg, and ggplot2. Make sure all are installed on your computer and load them at the beginning of your R script.*

**Complete the following questions/tasks:**

**Task I: Fit a model assuming constant occupancy and detection**

To complete this task, first read all three CSV files into R. Run some summary statistics to get a feel for what data you have. This may involve using the summary() function, the table() function, the head() (or tail()) function, or specific functions such as mean(), sd(), min(), and/or max().

Next, you will want to *z*-score standardize your continuous covariates using the *scale*() function. Hint, the *z*-score standardize the entire DOY data frame you can use the following code:

DOY <- as.matrix(DOY)

DOY[,] <- (DOY - mean(DOY))/sd(DOY)

head(DOY)

Next, you will need to combine the information within these three CSV files into a special type of object that unmarked needs to fit an occupancy model. This is the *unmarkedFrameOccu* object. You can use the following code:

bobcat\_umf <- unmarkedFrameOccu(y = EH, # Encounter history, must be a data frame or matrix

siteCovs = data.frame(zForest=site\_covs$zForest, zGrass=site\_covs$zGrass),

obsCovs = list(DOY = DOY)) # This must be a list of data frames, one for each covariate

*Note: You may notice that siteCovs uses columns with a “z” in front. Be sure you have done your z-score standardization!*

Finally, we will fit a model with constant occupancy and detection. This model assumes that occupancy and detection are each the outcome of their respective coin flip and the probability of flipping “1” is a single value (a single probability). Another way to think of this is that we are describing variation in our data (our 0’s and 1’s) using only the mean (remember that the mean of 0’s and 1’s is the proportion of 1’s or the probability of getting a 1). This model can therefore be thought of as an intercept-only model. This model is also often referred to as the “dot” model because the model may be expressed as: *p*(.)ψ(.)

Answer the following questions:

1. What is the proportion of sites where bobcat where detected?
2. What is the probability of use for the dot model?
3. What is the probability of detection for the dot model?

**Task II: Determine how forest and grassland land cover influence the probability that a bobcat used a site**

For this task you will fit some additional occupancy models using the proportion of forest and grassland land covers as covariates for occupancy. Remember that in the context of camera trapping, the occupancy probability is actually probability of use. You can use either a constant detection probability (i.e., *p*(.)) or let detection probability vary as a function of day-of-year (your choice!). Fit the following four models with their specified terms for occupancy:

* Constant occupancy
* Occupancy varies as a function of forest cover
* Occupancy varies as a function of grassland cover
* Occupancy varies as a function of forest **and** grassland cover

*Note: Don’t forget to use your z-score standardized covariates!*

Once you fit these four models you will need a means of empirically comparing the models to determine which one is “best.” What “best” means depends in part on the context of our analysis (e.g., is our focus primarily on generating unbiased parameter estimates or creating a model that makes accurate predictions). But for this exercise we will use a metric called Akaike’s Information Criterion or AIC. Actually, we will use AICc which is a variant of AIC that performs better than AIC when sample sizes are small. AIC is a widely used (and sometimes misused) metric for comparing models. Models with smaller AIC values are considered to have more empirical support and better fit your data.

Use the *AICc*() function within the AICcmodavg package to determine which of these four models has the most empirical support. For example:

AICc(name\_of\_your\_occu\_object)

Next, use the *summary*() function to look at the coefficient estimates (slopes) of the covariates for forest and/or grassland. Verbally describe the relationship between the probability that a bobcat will use the area around one of our cameras and one or both of these covariates.

**Task III: Plot the predicted relationship between probability of use and grassland and forest land cover**

For this task, you will take two of the models you fit for Task II modeling bobcat occupancy (the model with only forest and the model with only grassland) and use them to create a plot showing the predicted (i.e., expected) probability of use as a function of each covariate.

Remember that predicting from a fitted model object (whether that is a *glm* model object or an *occu* model object) requires that you provide a data frame (the *newdata =* argument within the *predict*() function) with the covariate values for which you need to make a prediction. This data frame must have column names identical to the column names of the covariates you used to fit your model. For example:

Occu\_mod <- occu(~1 ~Elevation, data = umf\_elevation)

New\_data <- data.frame(Elevation = …)

Remember that for prediction we would create a sequence of covariate values spanning the range of our actual or observed covariate. For example:

New\_data <- data.frame(Elevation = seq(min(umf\_elevation@siteCovs$Elevation),max(umf\_elevation@siteCovs$Elevation),length.out = 30)

*Hint: Have you z-score standardized your covariate?*

Once you have created your data frame you can use the following code as a basis for calculating your predicted values:

bobcat\_Forest\_pred <- predict(bobcat\_forest\_model,

type = 'state',

newdata = your\_new\_data\_frame\_you\_just\_created,

appendData=TRUE)

# Create a new column to denote that these are the predicted probabilities as a function of forest

bobcat\_Forest\_pred$Feature <- "Forest"

# This will help later on

colnames(bobcat\_Forest\_pred)[5] <- "Covariate"

Modify this same code to create an object “bobcat\_Grass\_pred” with the predicted probabilities as a function of grassland. Add a “Feature” column but use “Grassland” instead of “Forest.”

You now have two data frames with predicted probabilities. Combine them and use the following code from the ggplot2 package to create a plot.

all\_pred <- rbind(bobcat\_Forest\_pred[,c(1,3,4,5,7)],

bobcat\_Grass\_pred[,c(1,3,4,5,7)])

bobcat\_plot <- ggplot(data=all\_pred,

aes(x=Covariate))+

geom\_ribbon(aes(ymin=lower,ymax=upper,

fill=Feature),alpha=0.25)+

geom\_line(aes(y=Predicted,colour=Feature),size=2)+

scale\_y\_continuous("Probability of Use (psi)",limits=c(0,1))+

xlab("Covariate")+

scale\_fill\_manual(values=c("green","blue"))+

scale\_colour\_manual(values=c("darkgreen","darkblue"))+

theme\_bw()

bobcat\_plot

Verbally describe and compare the relationships between probability of use and these covariates. Include your graph in the report.

**Bonus Question (5 points):**

Create the same predicted plot as above but do this for detection rate and plot the probability of detection as a function of day-of-year (DOY).

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 8, 2024.**