

# Practical - Generalized linear model

*Biodiversity modelling*

*F.G. Blanchet – August 19–23, 2019*

## Introduction

This practical document presents R code for the section of the course regarding generalized linear model.

In the present document, we will focus on the distribution of *Acanthis flammea*.

## Load R packages

```
library(raster)
```

```
## Loading required package: sp
```

## Load the data

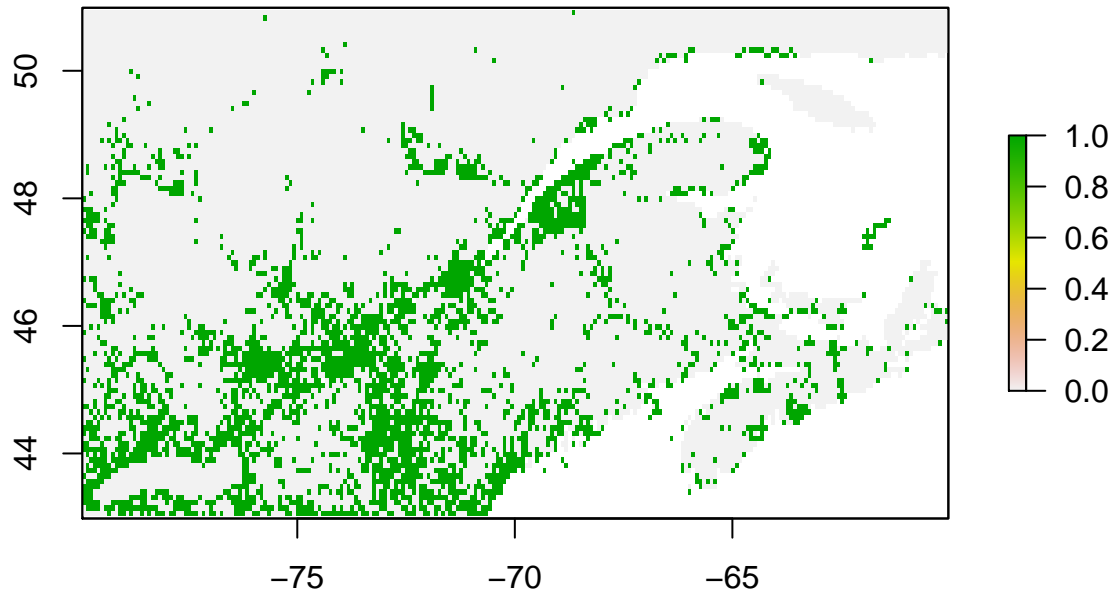
```
bird <- readRDS("birdAll.RDS")  
climatePresent <- readRDS("climate_Present.RDS")  
road <- readRDS("road_Distance.RDS")
```

## A bit of data visualization

It is always a good idea to look at the data a little before you start building models.

*Acanthis flammea*

```
sp <- bird$Acanthis.flammea  
plot(sp)
```



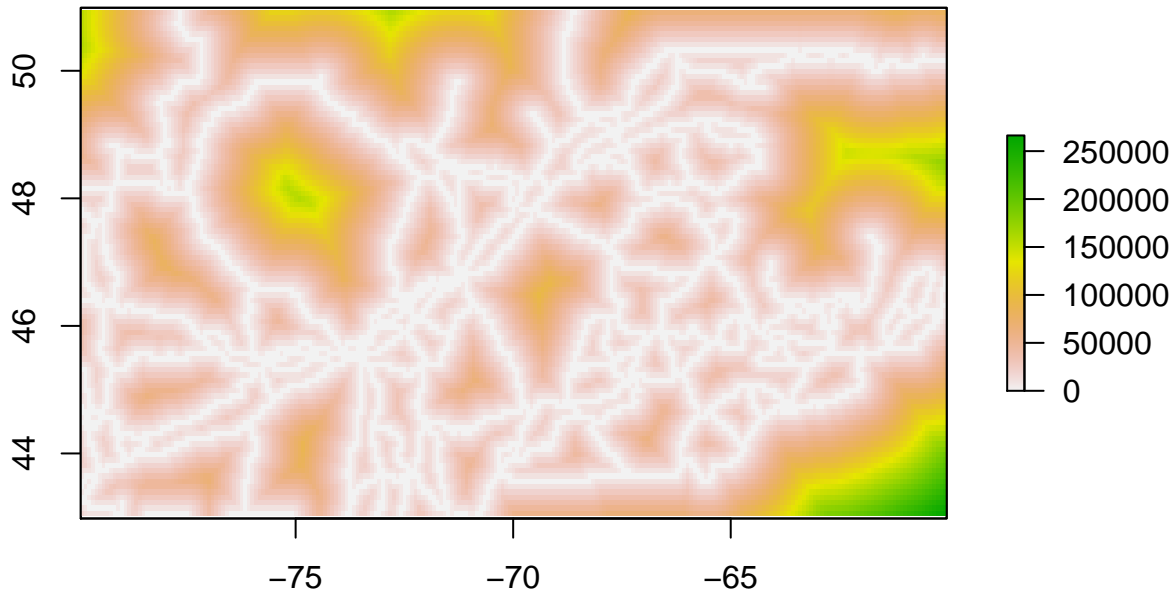
## Climate

```
climatePresent
```

```
## class      : RasterBrick
## dimensions  : 95, 239, 22705, 19  (nrow, ncol, ncell, nlayers)
## resolution  : 0.08327347, 0.08327347  (x, y)
## extent     : -79.93828, -60.03592, 43.02874, 50.93972  (xmin, xmax, ymin, ymax)
## crs        : +proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs +towgs84=0,0,0
## source     : memory
## names      :      bio1,      bio2,      bio3,      bio4,      bio5,      bio6
## min values  :   -2.046529,    5.570604,   20.333612,   669.529252,   17.693247,  -27.123456
## max values  :    9.7451844,   12.8244620,   30.8781696, 1291.6412141,   29.1501457,  -3.739012
```

## Distance to major road

```
plot(road)
```



## Build the model

Let's first start building a logistic regression for *Acanthis flammea* that includes only the climatic variable.

To do this, we must first extract the data from the raster files.

```
spDat <- values(sp)
climateAll <- values(climatePresent)
```

Because climate variables use different units, it is good practice to scale them (center and divide by their standard deviation) before using them in a model.

```
climateDat <- scale(climateAll)
```

Also, it is useful for the climate data to be in a `data.frame`

```
climateDat <- as.data.frame(climateDat)
```

Now we can perform a logistic regression

```
spClimate <- glm(spDat ~ ., data = climateDat,
                 family = binomial(link = "logit"))
```

Studying the regression parameters can be very informative to understand the environmental component that structure *Acanthis flammea*

```
summary(spClimate)
```

```
##
## Call:
## glm(formula = spDat ~ ., family = binomial(link = "logit"), data = climateDat)
##
## Deviance Residuals:
```

```
##      Min      1Q   Median      3Q      Max
## -1.5981 -0.6486 -0.3455 -0.1494  3.4230
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.031e+00  3.169e-02 -64.104 < 2e-16 ***
## bio1         1.382e+01  8.570e-01  16.122 < 2e-16 ***
## bio2         4.822e+00  7.394e-01   6.521 6.99e-11 ***
## bio3        -3.390e+00  4.534e-01  -7.478 7.54e-14 ***
## bio4        -2.502e+00  2.138e+00  -1.170 0.242025
## bio5         1.435e+05  1.680e+05   0.854 0.393166
## bio6        -3.469e+05  4.062e+05  -0.854 0.393181
## bio7        -3.024e+05  3.542e+05  -0.854 0.393166
## bio8        -3.499e-01  6.806e-02  -5.141 2.74e-07 ***
## bio9        -7.491e-01  1.217e-01  -6.156 7.46e-10 ***
## bio10       -4.943e+00  1.379e+00  -3.586 0.000336 ***
## bio11       -1.693e+01  2.832e+00  -5.977 2.28e-09 ***
## bio12        2.338e+00  4.608e-01   5.074 3.90e-07 ***
## bio13       -2.110e+00  2.470e-01  -8.540 < 2e-16 ***
## bio14        2.816e-01  1.679e-01   1.678 0.093434 .
## bio15        1.165e+00  1.817e-01   6.410 1.46e-10 ***
## bio16       -3.018e-01  2.845e-01  -1.061 0.288714
## bio17       -3.242e-01  3.799e-01  -0.853 0.393416
## bio18        2.356e-01  1.135e-01   2.075 0.037947 *
## bio19        6.379e-01  2.057e-01   3.100 0.001934 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 16541  on 17205  degrees of freedom
## Residual deviance: 13172  on 17186  degrees of freedom
## (5499 observations deleted due to missingness)
## AIC: 13212
##
## Number of Fisher Scoring iterations: 6
```

It may also be useful to project the model on map.

Two steps are necessary to do this task.

1. Estimate the model over the entire survey area
2. Project the estimation on a map

## Estimate the model over the entire survey area

```
spClimatePred <- predict(spClimate,
                        newdata = climateDat,
```

```
type = "response")
```

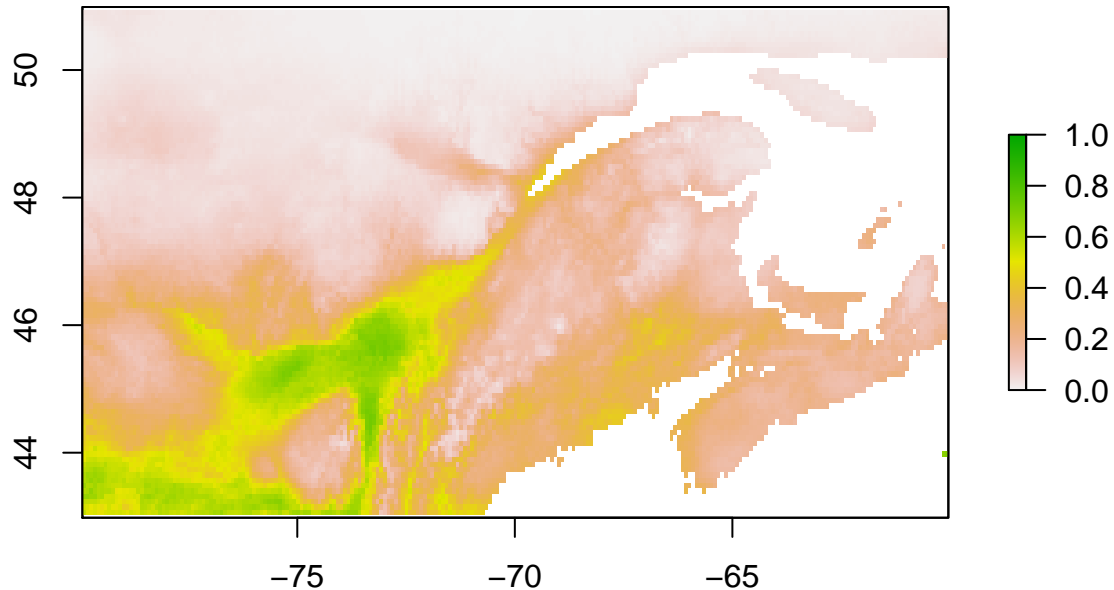
## Project the estimation on a map

This amounts essentially to formatting the data for it to be nicely projected. To do this, we will have to delve a little bit into the `raster` package.

```
# Build raster  
spClimateRaster <- raster(climatePresent)  
  
# Find where to place the model values in the raster  
values(spClimateRaster) <- spClimatePred
```

Now that the values have been organized, we can draw the map

```
plot(spClimateRaster, zlim = c(0,1))
```



## Question

Does this result make sense?

Can we do better ecologically speaking? The short answer is : Probably... Let's see if using *only* distance to major roads gives a good model.

```
# As before we scale the distance to road data  
roadDat <- as.data.frame(scale(values(road)))  
  
spRoad <- glm(spDat ~ ., data = roadDat,  
              family = binomial(link = "logit"))
```

Let's study the results

```
summary(spRoad)
```

```
##
## Call:
## glm(formula = spDat ~ ., family = binomial(link = "logit"), data = roadDat)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.9631  -0.7616  -0.4723  -0.1194   4.0333
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.29850    0.03954  -58.13   <2e-16 ***
## V1          -1.94257    0.05741  -33.84   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 16679  on 17246  degrees of freedom
## Residual deviance: 14596  on 17245  degrees of freedom
## (5458 observations deleted due to missingness)
## AIC: 14600
##
## Number of Fisher Scoring iterations: 6
```

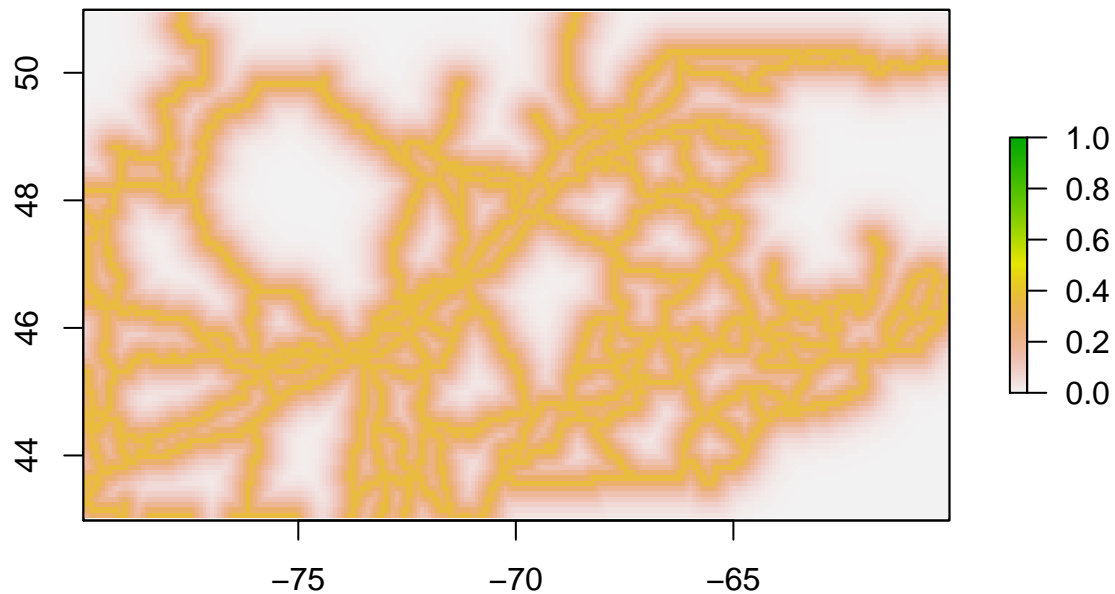
Now, let's draw a map of this result to better visualise the importance of the road for this species.

```
# Make prediction
spRoadPred <- predict(spRoad,
                      newdata = roadDat,
                      type = "response")

# Build raster
spRoadRaster <- raster(climatePresent)

# Find where to place the model values in the raster
locPix <- as.numeric(names(spRoadPred))
values(spRoadRaster)[locPix] <- spRoadPred

# Draw a map
plot(spRoadRaster, zlim = c(0,1))
```



For this species and actually for the whole bird data (because of the way the data was gathered), it makes sense that distance to major roads is so important. So, let's use only pixels within 50 km of major roads

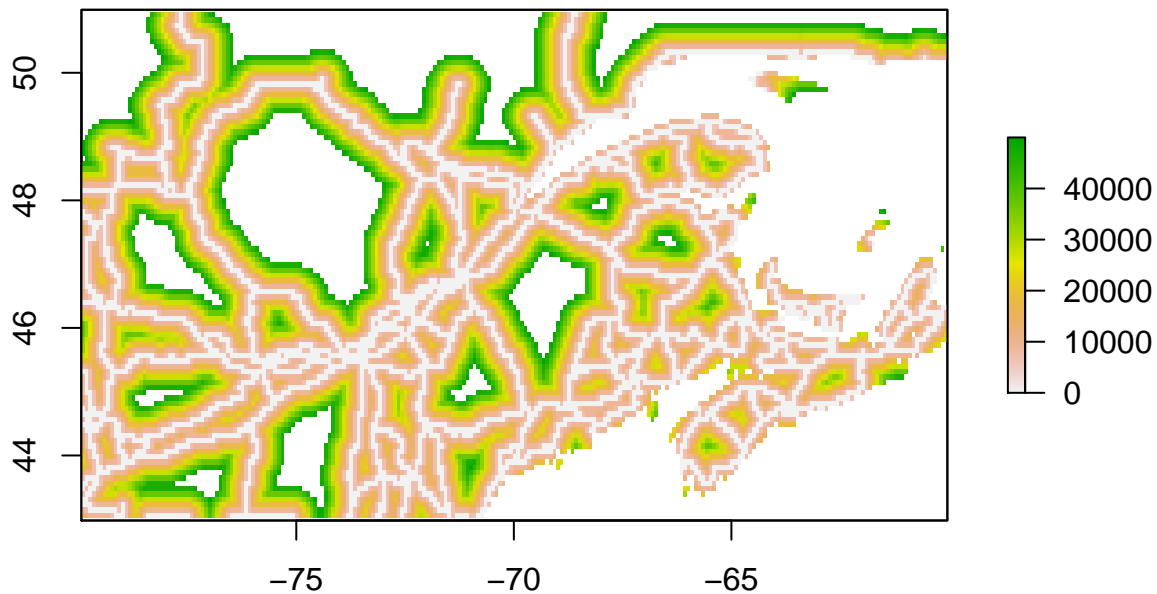
```
roadDat <- values(road)
road50 <- which(roadDat < 50000)
```

Geographically, we will focus on a subset of the full study area to estimate our models.

```
# Build the raster for the subset region
roadSub <- raster(road)
values(roadSub)[road50] <- values(road)[road50]

# Make sure only land pixels are considered
roadSub <- mask(roadSub, climatePresent[[1]])

# Draw the map
plot(roadSub)
```



## Logistic regression while trying to control for sampling bias

In this model we will take the distance to major roads into account by focusing only the pixels that are within 50 km of major roads.

Let's first isolate the pixels of interest to build the model and format the data for it to be used within our model.

```
# Find pixels with values
locPixRoad <- which(!is.na(values(roadSub)))

# For the species
spSub <- spDat[locPixRoad]

# For the climate
climateSub <- climateDat[locPixRoad,]
colnames(climateSub) <- colnames(climateDat)
climateSub <- as.data.frame(climateSub)

spClimateRoad <- glm(spSub ~ ., data = climateSub,
                     family = binomial(link = "logit"))
```

We can then study the results

```
summary(spClimateRoad)

##
## Call:
## glm(formula = spSub ~ ., family = binomial(link = "logit"), data = climateSub)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
```



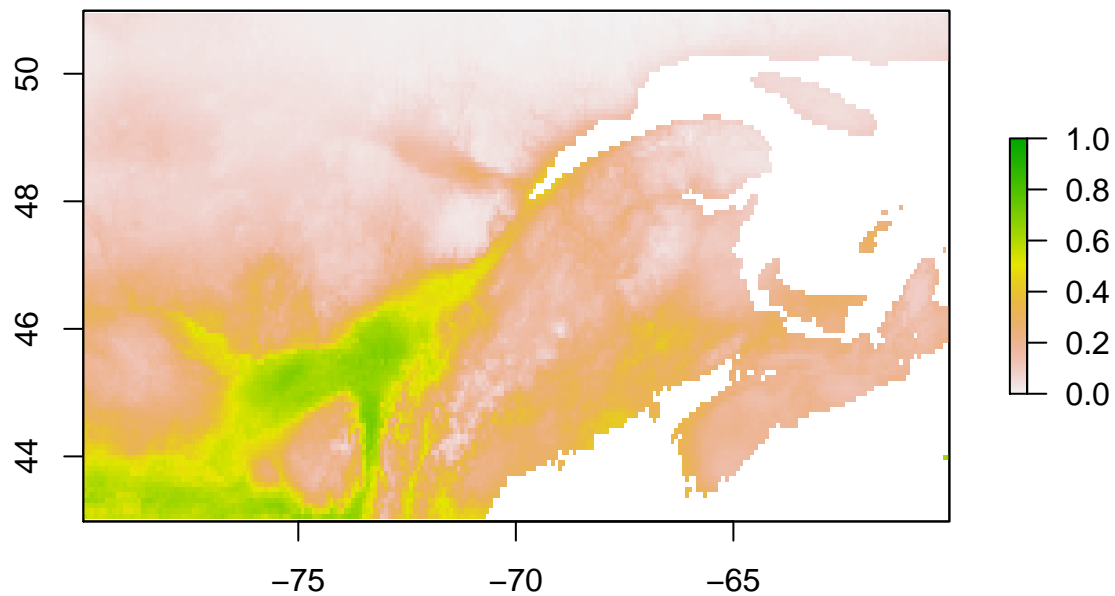
```
## -1.5867 -0.7184 -0.4474 -0.1902 3.2678
##
## Coefficients:
##      Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.851e+00 3.357e-02 -55.121 < 2e-16 ***
## bio1        1.310e+01 8.847e-01 14.804 < 2e-16 ***
## bio2        5.264e+00 7.433e-01 7.083 1.42e-12 ***
## bio3       -3.532e+00 4.550e-01 -7.764 8.20e-15 ***
## bio4        1.175e+00 2.208e+00 0.532 0.5944
## bio5        1.164e+05 1.717e+05 0.678 0.4977
## bio6       -2.815e+05 4.152e+05 -0.678 0.4977
## bio7       -2.455e+05 3.620e+05 -0.678 0.4977
## bio8       -2.234e-01 6.994e-02 -3.194 0.0014 **
## bio9       -6.013e-01 1.273e-01 -4.724 2.31e-06 ***
## bio10      -6.245e+00 1.407e+00 -4.438 9.09e-06 ***
## bio11      -1.173e+01 2.935e+00 -3.997 6.41e-05 ***
## bio12       2.431e+00 4.656e-01 5.221 1.78e-07 ***
## bio13      -1.993e+00 2.525e-01 -7.891 3.01e-15 ***
## bio14       1.666e-01 1.704e-01 0.978 0.3283
## bio15       9.333e-01 1.911e-01 4.883 1.04e-06 ***
## bio16      -2.163e-01 2.880e-01 -0.751 0.4526
## bio17      -4.659e-01 3.870e-01 -1.204 0.2285
## bio18       1.339e-01 1.157e-01 1.157 0.2473
## bio19       5.241e-01 2.086e-01 2.512 0.0120 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 14792  on 13815  degrees of freedom
## Residual deviance: 12398  on 13796  degrees of freedom
## (489 observations deleted due to missingness)
## AIC: 12438
##
## Number of Fisher Scoring iterations: 5
```

and project the model estimate on map.

```
# Prediction
spClimateRoadPred <- predict(spClimateRoad,
                             newdata = as.data.frame(climateDat),
                             type = "response")

# Build raster
spClimateRoadRaster <- raster(climatePresent)
values(spClimateRoadRaster) <- spClimateRoadPred

# Draw the map
plot(spClimateRoadRaster, zlim = c(0,1))
```



### A few things to try

- Model the distribution for another bird species
- Model the distribution for all bird species
- Model the distribution for a proposed climatic scenario

### To keep in mind

- This is not only model fitting, have an ecological perspective
- The data has particularities, be aware of it
- Remember the assumptions that you make when building your model
- Remember that the reason why you are modelling a species may skew the way you build and interpret the model.