

Practical - Generalized additive 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 additive 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")
```

Extract and organize the data

Let's organize the data based on what we learned from logistic regression.

```
# Species data
sp <- bird$Acanthis.flammea
spDat <- values(sp)

# Climate data
climateAll <- values(climatePresent)
climateDat <- scale(climateAll)
climateDat <- as.data.frame(climateDat)

# Pixels within 50 km of roads
roadDat <- values(road)
road50 <- which(roadDat < 50000)

# 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]])

# 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)

```

Build the model

To perform generalized additive model you need to load the R package `mgcv`.

```
library(mgcv)
```

Generalized additive models can be very useful to account for the non-linearity of space in the response variable.

To do this tensor splines are required and since it is not clear how our species of interest *Acanthis flammea* reacts to spatial forcing, let's use penalisation when building our model.

```

# Extract coordinates
xyAll <- coordinates(sp)
xySub <- xyAll[locPixRoad,]

explanaAll <- as.data.frame(cbind(climateDat, xyAll))
explanaSub <- explanaAll[locPixRoad,]

# Build formula
formu <- as.formula(paste0("spSub ~ ",
                           paste0("bio",1:19, collapse = "+"),
                           "+ s(x,y, bs = 'tp')"))

# Model estimation
spGAM <- gam(formu,
             data = explanaSub,
             family = binomial(link = "logit"))

```

Let's now study the regression parameters to understand the environmental component that structure *Acanthis flammea*. Do they differ from the logistic regression? How?

summary(spGAM)

```
##
## Family: binomial
## Link function: logit
##
## Formula:
## spSub ~ bio1 + bio2 + bio3 + bio4 + bio5 + bio6 + bio7 + bio8 +
##      bio9 + bio10 + bio11 + bio12 + bio13 + bio14 + bio15 + bio16 +
##      bio17 + bio18 + bio19 + s(x, y, bs = "tp")
##
## Parametric coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.289e+00  7.766e-02 -29.472  < 2e-16 ***
## bio1         2.538e+01  2.442e+00  10.394  < 2e-16 ***
## bio2        -2.459e+00  1.549e+00  -1.587  0.11240
## bio3        -5.548e-02  9.381e-01  -0.059  0.95284
## bio4         3.184e+01  5.421e+00   5.873  4.28e-09 ***
## bio5         1.286e+05  1.797e+05   0.716  0.47423
## bio6        -3.109e+05  4.345e+05  -0.716  0.47425
## bio7        -2.710e+05  3.788e+05  -0.716  0.47426
## bio8         4.371e-01  1.116e-01   3.917  8.96e-05 ***
## bio9        -5.252e-01  1.790e-01  -2.934  0.00335 **
## bio10       -3.872e+01  3.638e+00 -10.645  < 2e-16 ***
## bio11        3.305e+01  8.050e+00   4.105  4.04e-05 ***
## bio12        1.382e+00  8.370e-01   1.652  0.09862 .
## bio13       -4.142e-01  4.621e-01  -0.896  0.37009
## bio14        5.320e-01  3.340e-01   1.593  0.11121
## bio15        1.484e+00  3.511e-01   4.227  2.37e-05 ***
## bio16       -1.582e+00  6.087e-01  -2.600  0.00933 **
## bio17        1.400e+00  5.338e-01   2.623  0.00872 **
## bio18       -5.981e-01  2.062e-01  -2.901  0.00372 **
## bio19        3.833e-01  4.632e-01   0.828  0.40790
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##              edf Ref.df Chi.sq p-value
## s(x,y) 28.49  28.98  924.4  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.242   Deviance explained = 23.8%
## UBRE = -0.17686   Scale est. = 1           n = 13816
```

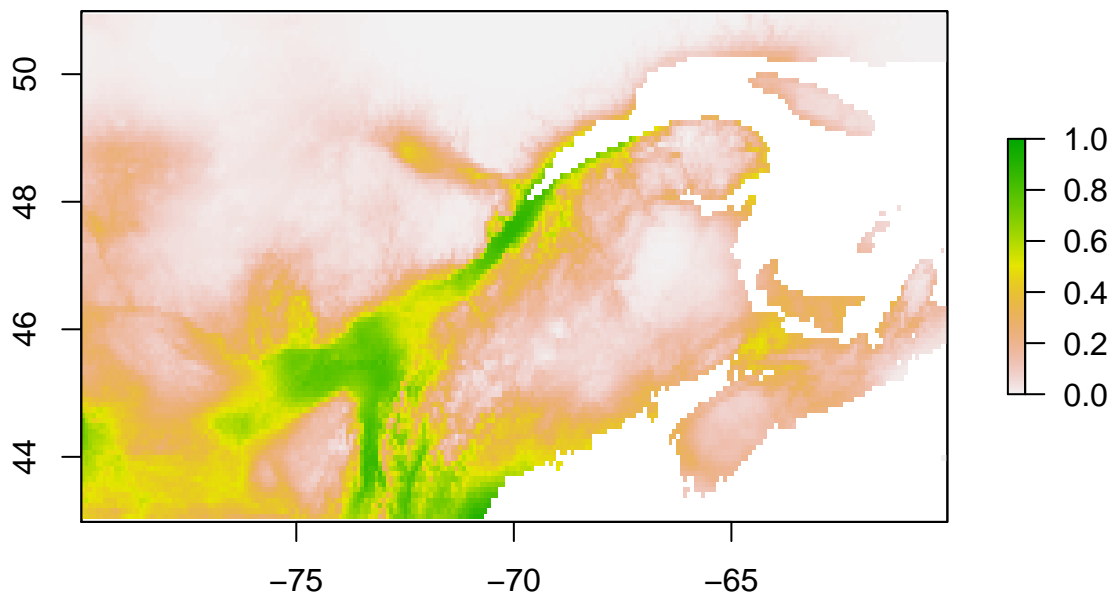
Project the estimation on a map

```
# Calculate estimated values
spGAMPred <- predict(spGAM,
                    newdata = explanaAll,
                    type = "response")

# Build raster
spGAM <- raster(climatePresent)

# Find where to place the values in the raster
values(spGAM) <- as.vector(spGAMPred)

# Map
plot(spGAM, 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
- Compare the result of a GAM with that of a GLM for the same species

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.