Generalized linear mixed models 2

Contents

Data

The aiv_ducks.csv dataset contains some of the data from the study by Papp et al. (2017) on the occurrence
of avian influenza (AIV) in populations of different species of ducks in eastern Canada.

> Papp, Z., Clark, R.G., Parmley, E.J., Leighton, F.A., Waldner, C., Soos, C. (2017) The ecology of
> avian influenza viruses in wild dabbling ducks (Anas spp.) in Canada. PLoS ONE 12: e0176297.
> https://doi.org/10.1371/journal.pone.0176297.

```
aiv <- read.csv("../donnees/aiv_ducks.csv")
str(aiv)
```

```
## 'data.frame':    8967 obs. of  10 variables:
##  $ Species          : chr  "MALL" "MALL" "MALL" "MALL" ...
##  $ Age              : chr  "HY" "HY" "HY" "AHY" ...
##  $ Sex              : chr  "M" "F" "F" "M" ...
##  $ AIV              : int  1 0 1 1 1 0 1 1 1 0 ...
##  $ Site             : chr  "Amherst Point" "White Birch" "White Birch" "Tower Goose" ...
##  $ Latitude         : num  45.8 46 46 46 46 ...
##  $ Longitude        : num  -64.2 -64.3 -64.3 -64.3 -64.3 ...
##  $ Year             : int  2005 2005 2005 2005 2005 2005 2005 2005 2005 2005 ...
##  $ Temperature      : num  18.6 17.6 17.6 17.6 17.6 ...
##  $ Population_Density: num  1.2 1.16 1.16 1.16 1.16 ...
```

Here is the description of the data fields:

- *Species*: Species code (ABDU = black duck, AGWT = green-winged teal, AMWI = American wigeon,
  BWTE = blue-winged teal, MALL = mallard, MBDH = black duck / mallard hybrid, NOPI = northern
  pintail)
- *Age*: Age (HY = hatching year, AHY = after hatching year)
- *Sex*: Sex (F/M)
- *AIV*: Presence (1) or absence (0) of avian influenza virus
- *Site*: Sampling site
- *Latitude* and *Longitude*: Geographical coordinates of the site
- *Year*: Year of sampling
- *Temperature*: Mean temperature in the 2 weeks prior to sampling
- *Population_Density*: Estimated duck population density (all species) for the site and year.

1. Fitting the model

a) Estimate the parameters of a comprehensive model to predict the presence/absence of AIV, including: the fixed effects of duck age and sex, temperature, and site population density; and the random effects of species, site, year, and site x year interaction (the latter is denoted as `(1 | Site:Year)` in the model). Should we check for overdispersion in this model?

b) What is the reason for including each of the random effects of the model in (a)? Check whether these random effects follow a normal distribution.

c) Produce a graph of the model-predicted probability of occurrence of AIV in (a) as a function of temperature for each of the four age and sex categories (HY/F, HY/M, AHY/F, AHY/M). The population density will not appear in the graph, but you can set it to its mean value for the predictions.

d) Starting from the full model, use the AIC to determine whether or not to include each of the following effects: temperature, population density, and the random effect for site x year interaction.

e) The authors of the original study determined a significant effect of population density by fitting a model with random site and year effects, but without their interaction. Why might the conclusions of your model differ from this result?

2. Model predictions

a) Add columns to the original dataset representing the prediction of the probability of occurrence of AIV (1) based only on the fixed effects of the model; (2) based on both fixed and random effects. Use the best model identified in the previous section.

b) For each type of prediction obtained (fixed effects; fixed and random effects), determine the predicted mean probability of occurrence of AIV for observations with AIV = 1 and the predicted mean probability of occurrence for observations with AIV = 0. Based on your results, do the model's fixed effects provide a good distinction between presence and absence? What about random effects?

c) Group the dataset by site and year, then calculate the mean longitude, latitude, and probability of AIV predicted by the full model for each site-year combination. Using these variables, produce a map of the sites with their predicted probability of AIV for each year. (You can use the facets in *ggplot2* to separate the graph into panels for each year).