Applied Analysis 2

Rohan Hore

August 2024

The data set for this analysis is taken from the paper *Shifts in timing and duration of breeding for 73 boreal bird species over four decades, H¨allfors et al, PNAS vol. 117, no. 31, pages 18557–18565 (2020)*

Experiment The data set consists of ringing records for birds across 73 species in Finland from year 1975–2017. Ringing is when a human volunteer places a metal ring on the leg of a bird living in the wild; these rings are marked and can be used to track the bird population. Ringing can only be performed on baby birds during a narrow window of age and is also dependent on conditions such as weather. This data set records the date, location, and bird species for all ringing events during the range of years studied. The dates of the ringing records can then be used to study the timing of the breeding season of the birds

Statistical analysis The paper studies how the breeding periods of the birds change with years. Specifically, the authors analyzed whether the breeding period is shifted/extended/shortened across species and locations, by using the ringing records of nestling in Finland from $1975 - 2017$. Ringing data have been shown to be a reliable estimate for breeding phenology. Nestlings can only be ringed when a certain size (which is species-specific), so the range of the ringing time is treated in the paper as a representation of the breeding period. For the ringing records of each bird species, bioclimate zone and year, **the authors calculated three metrics: the 5% and 95% sample quantiles, and the difference between them.** Denoting one of the metrics as y_{ijt} where i is a bioclimate zone, j is a species and t is a year. Then the paper proposes the following model

$$y_{ijs} = \beta_{js0} + \beta_{js1}x_i + u_{ij,s} + \epsilon_{ijs}$$

where x_i is the i-th year, and β_{js0} and β_{js1} are species-zone-specific coefficients. To borrow information across species, the authors added a prior on the βs, which takes the phylogenetic covariance matrix into account. The random effects $u_{ij,s}$ account for the correlations among species, and the authors assumed that $u_{ij,} \sim MVN(0, \Omega)$ where Ω is assumed to be a low-rank unknown covariance matrix. The random error terms $\epsilon_{ijs} \sim N(0, \sigma^2)$ are independent entries. This hierarchical model is solved by MCMC.

After obtaining the posteriors of β_{js1}, the authors found that many of the coefficients β_{js1} for the two quantiles are significantly negative, uncovering a general advance of breeding with a strong phylogenetic signal but no systematic variation over space. For some species, the difference of the quantiles also have a negative trend, indicating that the breeding periods are more concentrated. In addition, these differences are more apparent in resident and short-distance migrating species that breed early in the season.

Data Reading

```
#data reading
ring_data=read.csv("73_species.csv")
traits_info=read.csv("Traits_73_species.csv")
#looking at data
head(ring_data)
head(traits_info)
```

```
##   NestID Species XEUREF  YEUREF BZ Day Month Year Dayofyear
## 1  A_001  ACCGEN 270000 6630000 HB  19     6 1979       170
## 2  A_002  ACCGEN 280000 6630000 HB  18     6 2001       169
## 3  A_003  ACCGEN 280000 6640000 HB   6     7 1984       188
## 4  A_004  ACCGEN 280000 6640000 HB  26     6 1987       177
## 5  A_005  ACCGEN 280000 6640000 HB  25     6 1995       176
## 6  A_006  ACCGEN 280000 6640000 HB  24     6 2010       175
```

```
##   Abbreviation      Scientific.name Broods Mig
## 1       PHACAR  Phalacrocorax carbo      1   S
## 2       ARDCIN        Ardea cinerea      1   S
## 3       SOMMOL Somateria mollissima      1   S
## 4       BUCCLA   Bucephala clangula      1   S
## 5       PERAPI      Pernis apivorus      1   L
## 6       CIRAER   Circus aeruginosus      1   L
```

1 Possible Questions

Problem 1

- Is i.i.d assumption of individual ringing events a reasonable one? Did the authors make this assumption?
- If we consider the above model (the one from paper) with upper/lower quantiles as our response, what aspects have been ignored in the model and how would you take care of it?

Problem 2

In Figure 2 of Hallfors et al, left-hand panel, the authors show eight estimated densities (4 species in each of 2 different years) of ringing events across the seasons. Several of these estimated densities are multi-modal.

Attempt to produce versions of these eight plots from the same data, improving them where possible.Comment on whether the published density estimates accurately represent the raw observed data. Do you think the multi-modal features in the published density estimates likely represent clustering in the actual nesting behavior of the birds? Are there other possible explanations?

Problem 3

In Supplementary Text $S2$, entitled "Assessing change in ringer behavior", the authors discuss whether ringer behavior and effort may change over time. In particular they examined changes in wing length (Fig 1) [the data for which are not included in the dataset we used here]. Why the authors examined changes in wing length over time?

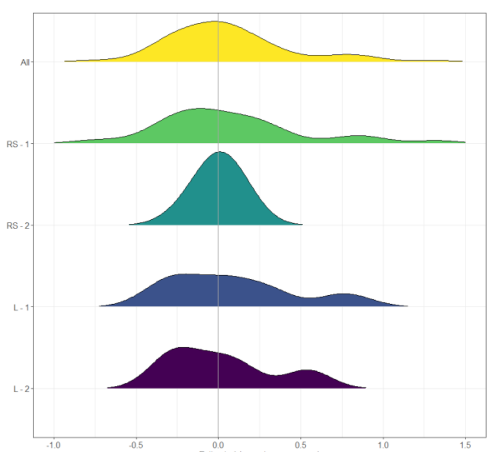


Figure 1: Figure S7 from supplement

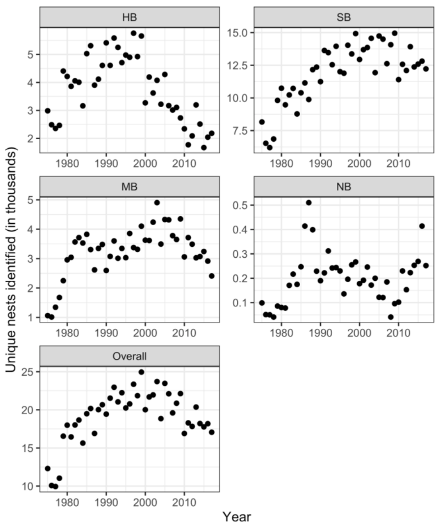
Problem 4

They also analyzed changes in total numbers of ringing events as a proxy to ringing effort (Figures 2a,2b) over the years of the study. Explain how changes in ringing effort over time, if they existed, could impact the main results and conclusions of the study. Authors argue

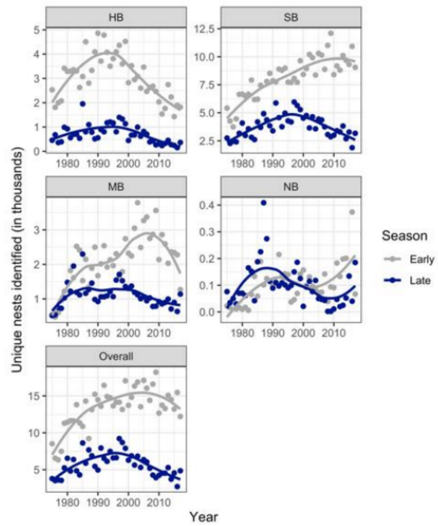
- Annual effort levels stabilized in all zones by 1990 and remained relatively constant thereafter. There are sufficient samples (in thousands) late in the breeding season to accurately identify the end of the breeding.

- While there is variation in the number of nests observed in each year over the study period, the proportions of nests observed in the beginning and end of the breeding season are relatively constant over the study period

Discuss the strengths and limitations of their argument. What assumptions would make it possible to distinguish changes in ringing effort from changes in bird behavior?



(a) Figure S9 from supplement



(b) Figure S10 from supplement

Figure 2: Supplement pictures

Problem 5

For this question, we will work with the duration of the breeding season as the response variable, which(as in the paper) is defined as the number of days between the 5th and 95th percentile of ringing events. For the species CORRAX, use a linear regression to assess whether the duration of the breeding season is showing a decrease in length over time. (For this problem, you should treat duration as the observed response, and ignore the fact that it was computed based on individual ringing events. However, you may choose to include other information such as boreal zone in your analysis.) Discuss your findings.

Problem 6

For this question, we will also work with the duration of the breeding season as the response variable, and consider the linear regression model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

where y_i is the estimated duration of the breeding season for year i and x_i is corresponding time variable for the i-th year (e.g., $x_i = 1990$). We focus on fitting this model for the species HIRRUS in the south boreal zone. The following histogram reveals that, for this species-zone combination, the number of observed ringing events in each year has a dramatic increase after year 1997. Explain why this fact might call into question the use of ordinary least squares (OLS) for fitting the model above. Propose and implement an alternative

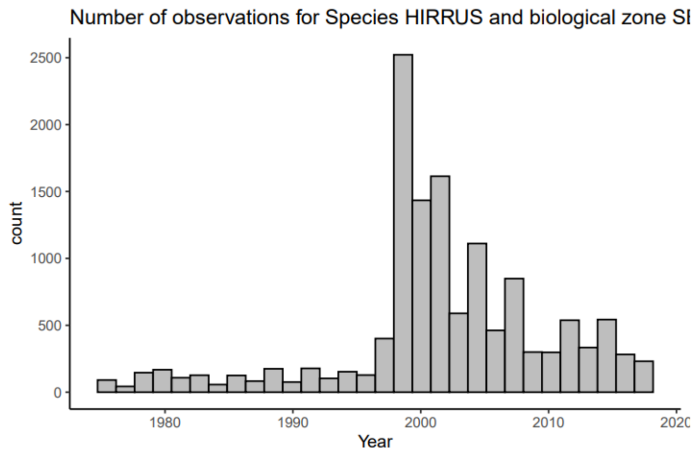


Figure 3: Number of ringing events per year for HIRRUS species

approach, and compare the results with those from OLS. In what scenarios might this kind of issue be a serious concern in practice?

Problem 7

In this question we consider an alternative way of measuring shifts in the breeding season than the quantile-based approaches used in the paper. Specifically, for each year i, let p_{ijs} denote the proportion of ringing events that occur before Day 170 of the year for species j in bioclimatic zone s. Changes in pijs with year would suggest shifts in the breeding season (for species j in zone s). Devise analyses to estimate how p_{ijs} changes with year i, and whether there are differences in this year effect among species. Further, assess whether year effects differ among species that have different numbers of broods or migration types. Clearly explain the models you use, the results of the analyses, and your conclusions.