447 Project Proposal

Tim 30407324

Github

Repo

TL;DR

We'll apply bayesian inference in the domain of species density estimation in ecology. Specifically, we'll estimate and predict the density of migratory birds and fish in Canada employing bayesian spatial models and time series models.

Datasets

- **Birds** – Birds count dataset is accessible through Breeding Bird Survey USGS BBS by querying by region for BC. It contains annual observations count data across more than 50 years for BC routes.
- **Fish** – fish count dataset is accessible through Government of Canada, containing location-clustered capture data of fish across Canadian lakes, streams, and oceans.

Here's a preview of the Fish dataset:

```
library(magrittr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
fish = read.csv("fish.csv") %>%
  select("Species", "SITEID", "Date","Waterbody.Name","WaterbodyType","Province")
head(fish)
```

```
##                 Species SITEID      Date  Waterbody.Name WaterbodyType Province
## 1 Micropterus salmoides      1 24-Sep-02 Ausable Channel        Stream       ON
## 2     Notropis anogenus      1 24-Sep-02 Ausable Channel        Stream       ON
## 3      Lepomis gibbosus      1 24-Sep-02 Ausable Channel        Stream       ON
## 4      Perca flavescens      1 24-Sep-02 Ausable Channel        Stream       ON
## 5     Lepomis peltastes      1 24-Sep-02 Ausable Channel        Stream       ON
## 6   Lepomis macrochirus      1 24-Sep-02 Ausable Channel        Stream       ON
```

Here's a preview of the Birds dataset.

```
bird = read.csv("bird.csv")
names(bird) = gsub("X", "Year", names(bird))
```

```r
bird = bird %>%
  select("Species.List", "Year1968", "Year1969","Year1970","Year2020","Year2021","Year2022")
head(bird)
```

```
##     Species.List Year1968 Year1969 Year1970 Year2020 Year2021 Year2022
## 1    Route Count       15       13       17        0       47       58
## 2 Cackling Goose        0        0        0        -        0        0
## 3   Canada Goose       12        0        0        -      175      315
## 4      Mute Swan        0        0        0        -        0        0
## 5 Trumpeter Swan        0        0        1        -       20        4
## 6    Tundra Swan        0        0        0        -        0        0
```

Project Themes

The two themes I'll aim to explore are spatial models and time series models, they are suitable because the bird dataset contains yearly , and the fish dataset provides clustered location, and these attributes are highly meaningful from ecological preservation perspectives.

Spatial Models

On the spatial model side, I'll follow Moraga Ch9 and start with testing out BYM model which assigns independent and non-identicle priors to latents from different regions.

Time series Models

For the Birds dataset, we'll employ a time-series model to infer the density of each bird species. I plan to start with experimenting with time series in two settings: prediction and fill in blank. The blank comes from the no data gap in 2020, in which case there's no ground truth to verify, but we could also set aside 2021-2022 (the most recent data) as prediction and see how the model does on those as a rough estimate of how accurate the estimate we have on 2020 is.

Team Contribution Assignment

N/A, single person group.