

Finding questions

Summary of previous plots

1. Flock size and number of families show a linear relationship on a log-log axis and are not very good predictors of juvenile percentage.
2. Both flock size, but more pronouncedly, juvenile percentage, show a non-linear relationship with time. At the inter-annual scale, both flock size and juvenile percentage follow the lemming cycle, being low in lemming crash years.
3. Intra-annually, flock size shows the expected trend: as more geese arrive over the winter, the flock sizes increase, and as they begin to leave or move eastwards in late winter, flock sizes decrease.
4. Juvenile percentage shows a different trend, and rises after mid-winter, being highest in late winter when breeding adults have moved east or begun migrating.
5. Early winter juvenile percentage shows more variation: in some winters, such as in 2002, the percentage of juveniles is higher in October than at any other time in the season, while in others, such as 2009, the percentage of juveniles increases monotonically from a low value in October to a high values in late winter.

Overall, juvenile percentage and family size show the most change and variation through the winter.

Questions

1. What drives juvenile percentage and family size?
2. Why is there variation in the early winter juvenile percentage?
3. Do successful families arrive earlier or later in some years?
4. What is a large flock? A large flock should be defined as the 0.75 quantile for total flock size for all flocks within a unit of time.
5. What are the peak migration times in autumn and spring for each year?

Response variables

1. Juvenile percentage: available from all regions/zones, easily calculated if missing.
2. Family size: mean family sizes available from ~1200 flocks which contain family size data until 2011. Raw size of each family counted, available from around 33000 families recorded over 1200 flocks.

Predictors

1. Time: Time can be calculated from the year/month/day field provided in the data. From a single POSIXct time variable, the year, month and day can be extracted. It can be manipulated to return the time since or to an arbitrary date (with midwinter defined as Jan 1st), or to a data driven date, such as the arrival and departure times of some percentage of geese. Variables to use: **time**, **year**, **time since arrival**, **time to departure**.
2. Zone: Provinces are classified into zones:
 - **Rhinelands**, includes Duisburg, Wesel, Cleves, Brabant, Limburg, Nijmegen, and Wageningen.
 - **IJsselmeer**, includes Frisia, Flevoland, and North Holland.
 - **Southwest**, includes Zeeland, and South Holland.
 - **East Frisia**, includes Leer, Drenthe, and Groningen,
 - **Other**, all other provinces.
3. **Coordinates**: Coordinates are geocoded from the field **Site_name** in the data, ie, an automated search queries the names of each unique value of **Site_name** via GoogleMaps and returns a pair of coordinates if found. For the 33000 families in the data, around 31000 have a valid, geocodable **Site_name**.
4. **Temperature, wind direction/number of days with x wind direction**, and **NDVI** (including water and snow): To be retrieved from online remote sensing data, monthly composites to be used. May be retrieved from weather station data. Families may spend more time on water, water data may have to be skipped.
5. Random factors:
 - **Year**
 - **Region**
 - **Food type**
 - **Flock size**