

Goose Flock Geocoding 01

March 27, 2017

Geocoding

1. Nearly all flocks were recorded as having juveniles, but the number of families was not recorded in all flocks (see previous document on data description). Neither is any explicit coordinate information given for any of the records. Here, I attempted to geocode the records, ie, assign coordinates using an automated online search.
2. I reduced the dataset to only those records taken in the Netherlands or North Rhine Westphalia. These were around 5100 of 5600 records, of which, I selected only those records after the year 2000.
3. I further reduced the records to those in which at least one family was recorded. While 5162 records had flock data, only 1246 were recorded as having at least one family.
4. These families were recorded at 376 unique sites, each of which was named. This is in contrast to the larger dataset containing flocks for which family counts were not available (still restricted to NL and NRW), in which flocks were recorded at possibly an additional 104 unnamed sites.
5. Of these 376 sites, geocoding using a GoogleMaps API succeeded for 244 of them.
6. These included sites where the geocoding did not correctly identify the location and produced NA values (these were invariably in pairs, ie, both lon-lat were NA) and sites where the geocoding produced coordinates that were not in the study sites defined above.
7. The coordinate boundaries of the study site were set between 0 - 8 degrees E and at 54 degrees N. After removing records of such values and NAs, 754 records remained.

Map: Flock size and mean family size

Map 1 shows mean family size (colour scale) and flock size (size scale) for the years 2000 - 2011 in the Netherlands, North Rhine Westphalia and Lower Saxony. Blue crosses represent the geocoded locations. Circles representing each flock have been jittered to make them visible, since there is a fair amount of overlap.

It isn't clear how points in Lower Saxony have been included in the study area, since the data was specifically restricted to the former two regions. The distribution of points in Lower Saxony also doesn't appear to be random, being concentrated in East Frisia, a known location for wintering geese and a base of goose counters. This could be an error in the data and may have to be accounted for by explicitly reducing the longitudinal extent of the points.

One possibility is that during mapping, the CRS of the geocoded points and that of the shapefile representing the border are not identical, causing points in the Netherlands to appear to be in German East Frisia. This issue was eliminated by explicitly setting the CRS of the data to be the same as that of the shapefiles. The standard WGS84 projection was used.

The remaining possibilities are that there were errors in the geocoding process by which points in the Netherlands with similar names to points in East Frisia were located there instead, or, that the data when initially recorded in the field included some East Frisian points under the Netherlands.

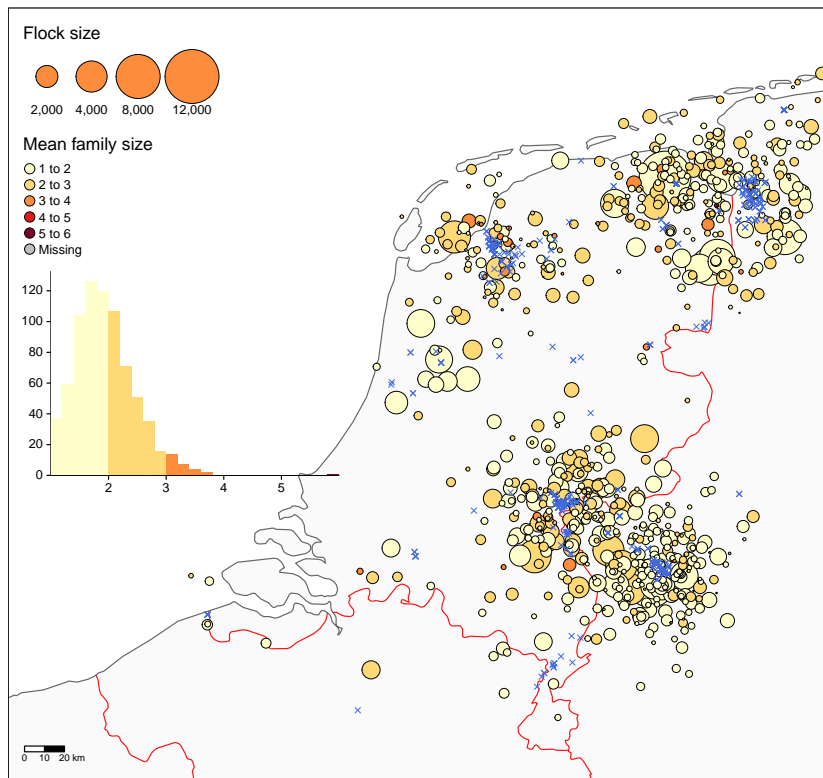


Figure 1: Map of mean family size per record on the wintering grounds, 2000 - 2011. Blue crosses represent geocoded locations of the records. Circles are jittered to make them visible.

Map: Flock size and juvenile percentage

Map 2 shows the flock size (size scale) and the percentage of juveniles (colour scale) in the same records as above. Blue crosses represent the coordinates of geocoded points.

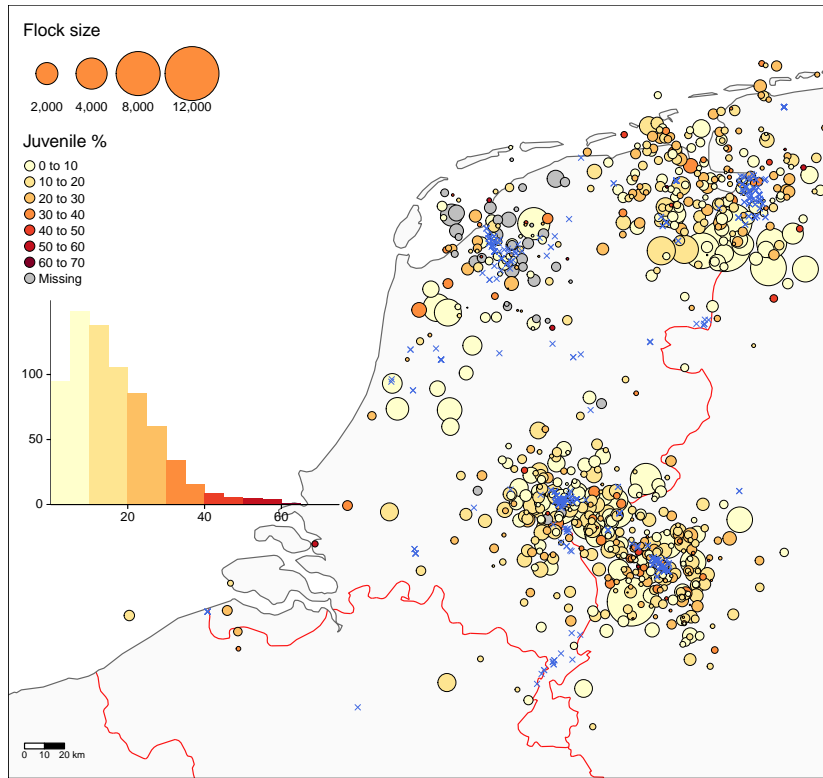


Figure 2: Map of juvenile percentage per record on the wintering grounds, 2000 - 2011. Blue crosses represent geocoded locations of the records. Circles are jittered to make them visible.

Mapping: Adding more data

1. While there are only around 1200 records with family size data sampled in the Netherlands, North Rhine Westphalia and Lower Saxony, most records do have data on the juvenile proportion – 5117 records with juvenile proportion out of 5162 records from the study site total.
2. Similarly, while only 754 records from the study site have geocodable family counts, 5117 records have juvenile proportions.
3. Geocoding on 376 unique sampling locations yielded only 244 coordinate pairs, greatly reducing the number of records, or flocks, included in the visualisation.

4. In an effort to utilise more of the data, I instead attempted to geocode for the data field Region, since this is likely to yield more specific locations than the field Country (which would of course yield only two). Within the study site, there are 22 unique regions.
5. These regions are noted as abbreviations, such as FL, FR, NB, OV, RG, NL. Using a table, it was possible to determine which region lies in which country.
6. Using a table of Netherlands province codes, I assigned names to each of these abbreviations. Ten more regions were noted than there are provinces in the Netherlands. Of these 10 extra, one, Leer, was easily identified, as were three more that lay in the Rhinelands: the cities of Cleves, Wesel and Duisburg. Of the remaining six, IJ was assigned as the river IJssel, NL as the Netherlands in general, WG was guessed at as being the city of Wageningen, RG was assigned as the city of Nijmegen, and UT as the city of Utrecht. LB was not found, however it has only one record. For convenience, the province of South Holland was assigned to Rotterdam, since it was causing the GoogleMaps API some confusion.

Map 3: Juvenile percentage by region

1. I visualised the data above on a map, as seen here, in order to look for any spatial patterns in the juvenile proportion of flocks, and in the flock size. River courses and lakes were added for context, even though the coordinates retrieved for each province are only representative.
2. The largest flocks appear to be from provinces close to the IJssel, the Waal, or the IJsselmeer. Three province means, from North Holland, Overijssel and Flevoland are high for both flock size as well as juvenile percentages.

Map 4: Juvenile percentage by region in each month

I then separated the data by month, to look for interactions between temporal and spatial patterns.

On the whole, as the winter progresses, juvenile proportions increase in flocks, and flocks appear to show an increase as well.

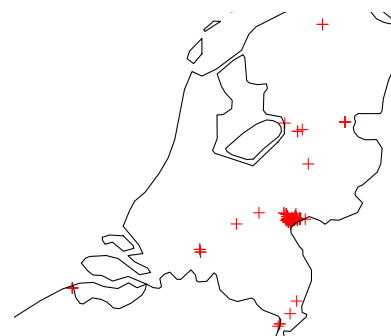


Figure 3: Records where the region is noted as RG lie closest to Nijmegen.

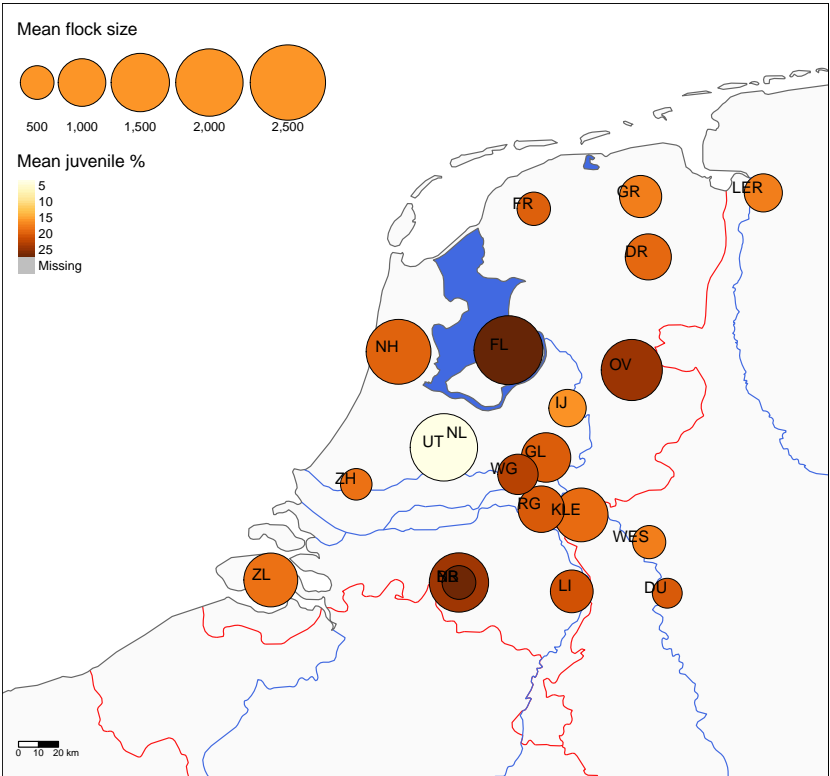


Figure 4: Mean juvenile percentages and flock sizes for each region noted in the data. The coordinates for the region are only representative.

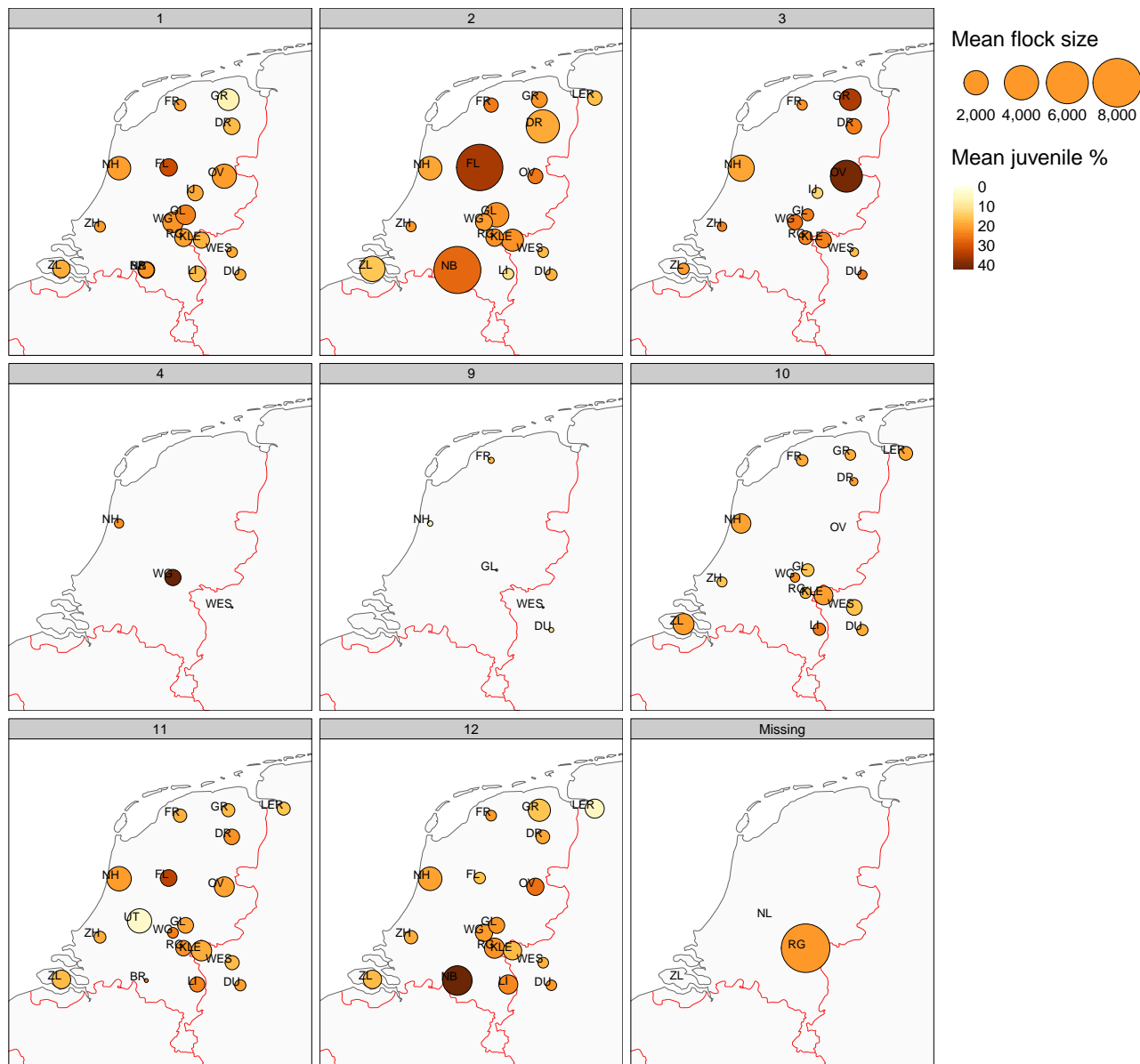


Figure 5: Mean juvenile percentages and flock sizes in each province in each month recorded in the data.