Goose family sizes from position loggers

The goal of the exercise was to see how the family sizes of position logged families changed over time. The immediate objective of the exercise was to calculate the distance between the first goose (in most families the adult male/female, but see the Wo family) and all other geese in the family. Subtracting two may be expected to give the number of juveniles present with their parent(s).

Thirteen families' tracks were provided as kml files with the flight bursts removed. Each family's tracks were included in a single file, with individuals in separate layers. The tracks were imported as SpatialPointsDataFrames in R, using the readOGR function for kml files. Each layer, or individual, formed a separate SPDF. The SPDFs corresponding to each family were stored in a list, and thirteen such lists made up a higher list.

The data contained only one relevant variable, the timestamp, stored under the header "Name". The coordinates for each timestamp were held in the coordinates element of the SPDF. No mismatches between timestamp and coordinate length were found, and none was expected.

For each individual, the timestamp was first extracted as a POSIXct object, and these timestamps were stored in a list of lists structured as above. All timestamps where the month was greater than 3, and less than 9, were removed. The intersection of the sets of timestamps of individuals of the same family is low, ie, two individuals from a family do not share many timestamps. This is because the POSIXct class accommodates timestamps which have a resolution of seconds, resulting in small differences between the times at which a position was logged. Each timestamp was rounded to the half hour, both to reduce the size of the data frame and the computational load, and since this is also the expected interval of the position loggers. To this list, the coordinates of each individual were added, forming a dataframe for each bird. Coordinates were averaged for every half hour in cases where the timestamps were duplicated. This was common in the data retrieved from the 2016 birds, which typically had twice the number of data points expected from a logger that records positions half hourly.

The thirteenth list's order was reversed because the kml file containing this family (Wo) had its layers in reverse order, with the adult male as the last layer. The first layer represented a juvenile goose which stopped transmitting after two

months. Following this, all individuals within each family were merged into a single family-wise dataframe with the timestamp as the variable to merge on.

The distance between the first individual and the others was calculated using the distGeo function from the package Geosphere. Instead of a distance matrix with distances between all pairs of points, the distance between each bird's position at each timestamp was calculated. This resulted in $i \cdot j$ distances, where i is the number of timestamps and j the number of individuals. This vector was turned into a matrix of j columns and filled by row. The process was judged to have succeeded since the first column of each 'distance matrix' had zeroes, as it held the distance from the first bird to itself. These matrices were turned into a dataframe and the timestamps added.

At this stage, the family size was calculated as the number of individuals within a radius of 1000m metres, with debatable results. Since the temporal scale of the study is around six months, recording family sizes every half hour may represent data that have a higher resolution than necessary. To counter this, the maximum family sixe of each family on each day was found from the half-hourly data. The results of this process were better than the half-hourly data and had fewer gaps due to missing values. However, the process did not improve the data for some families, especially those which lost juveniles (Wo, Ev) or which disbanded (Ev, Jo?).

The entire process was repeated considering only timestamps between 20:00 and 06:00 hours, since this was expected to represent the roosting period during which geese are less active, and frequent position changes should ideally not have resulted in errors from position averaging. The results were the same.