

Methods: Analyses

Contents

Hypotheses and predictor variables	1
Hypotheses	1
Predictor variables	2
Hypothesis testing	3
Modelling family sizes	3
Modelling flock sizes and juvenile proportions	4
References	4

Hypotheses and predictor variables

Hypotheses

We sought to test the following three hypotheses related to family size: 1. More successful families winter in the west, 2. Larger families are seen in smaller flocks, and, 3. Family sizes decrease over the winter. To disentangle the relationship between flock size, family size and space, we also tested whether 4. Flock sizes are smaller in the west. Finally, we also tested whether 5. The proportion of juveniles in flocks increases over the winter.

Predictor variables

We chose main predictors to match our hypotheses. The longitudinal position of each observation, the flock size, and the time in winter were selected as hypothesis testing predictors, and were always included in further analyses. We added the summer predation index as an auxiliary predictor. Longitude was available for all records in both the observations of marked geese and the flock counts. While the individual observations were expected to hold information on the number of flock members and the habitat type, we could not obtain these and the analyses could not include flock size as a predictor when modelling the individual observation data.

We then chose between time since arrival, and time to departure to represent the time in winter as experienced by geese. To do so, we specified two generalised linear mixed models (GLMMs) from the `lme4` package (Bates et al. 2015), with one model using the number of days since the first goose arrivals on the autumn migration, and the other using the number of days to the end of goose departures on the spring migration as predictors of family size. We ran these models on the family level data, and included breeding year, observer identity, and habitat type as *iid.* random effects. We ranked the two models by their AICc scores (here equivalent to AIC) using the MuMIn package (Barton 2016), and days since arrival ($AICc = 25,396$) was selected over days to departure ($AICc = 25,729.7$) as a proxy for the time in winter.

To quantify summer predation, we first obtained rodent (lemming) abundance indices for the breeding grounds from the website *Arctic Birds* (www.arcticbirds.net), an initiative of the International Breeding Conditions Survey on Arctic Birds. Nolet et al. (2013) previously used the same approach to fill in gaps in the dataset they used. Sites on *Arctic Birds* are marked on a map and binned into four classes: “Unclear”, “Low or absent”, “Average”, and “High”. The same sites are not present in each year. We graded these sites on a 0 – 3 scale, with 0 for “Unclear, and 3 for “High”. The distinction between sites graded 0 and 1 was itself unclear. Both 0 and 1 were used in different years to mark the island of Kolguyev, which is well known to have no lemming cycle. We took 0 to indicate a near or full absence of lemmings rather than an unsurveyed site, and also included an entry of 0 for Kolguyev in each year. Breeding birds form the major prey base for the island’s predators, the majority of geese on the island are breeders (Kondratyev and Zaynagutdinova

2008, Kruckenberg et al. (2008)), lending inclusion of the index for Kolguyev quite some value. We averaged the lemming index across the sites in each year, and then for each year t , we calculated a predation index (P_t).

$$P_t = \frac{L_{t-1} - L_t + 3}{2} \quad (1)$$

We matched this predation index to observations by the breeding year.

Hypothesis testing

Modelling family sizes

We used the family data and the individual data to model family sizes and so test the first three hypotheses. We constructed generalised additive models (GAMs) using the `mgcv` package (Wood 2011). GAMs are generalised linear models which allow the inclusion of smoothed terms, where the response has a linear relationship with a smooth function of the predictor variable (Wood 2006). As implemented through the `mgcv` package, they allow for the modelling of simple random effects through the use of a random effect smoothing basis. This treats the random effects as penalised smooth terms, and within the R environment, is held to be faster and numerically more reliable than the modelling of these effects in generalised additive mixed models (GAMMs) (Wood 2013).

The main predictors were included as fixed parametric effects. We further specified the breeding year, the habitat type, and the observer as random effects in the model run on family data. We could not retrieve the habitat type and observer identity for the individual observations, and these could not be included as random effects when modelling the individual data. However, we did have information on the individual identity, and this was included as a random effect. Table 2 shows the models and their effects. In both cases, the random effects were modelled as “re” smoothed terms, and assumed *iid*. We used data for which there were complete cases for each and every one of the fixed and random effects. Since family sizes are discrete counts that can only take whole

number values and are bounded at the lower limit by zero, we specified a Poisson error distribution for our model.

Modelling flock sizes and juvenile proportions

We used the flock count data to model the flock sizes and the juvenile proportions, constructing separate GAMs for each. We included each of the main predictors in both models, except flock size when it was the response. When modelling juvenile proportions, we made two changes to the model structure. First, we used a binomial error distribution as is appropriate for proportion data. Second, we specified the number of days since goose arrivals as a smoothed term using a thin plate spline penalty basis (Wood 2003). We specified a maximum of four knots, allowing the GAM to choose up to that number using generalised cross validation (Wood 2004). We specified the same *iid.* random effects in both models as for the family data GAM above and in the same way.

Table 1: Model structures; effect codes: *1* Longitude, *2* Flock size, *3* Days since arrivals, *4* Summer predation index, *5* Breeding year, *6* Habitat type, *7* Observer identity, *8* Goose identity.

Response	Data source	Fixed effects	Random effects
Family size	Family counts	1, 2, 3, 4	5, 6, 7
Family size	Individual observations	1, 3, 4	5, 8
Flock size	Flock counts	1, 3, 4	5, 6, 7
Juvenile proportion	Flock counts	1, 2, 3, 4	5, 6, 7

References

- Barton, Kamil. 2016. *MuMIn: Multi-Model Inference*. <https://CRAN.R-project.org/package=MuMIn>.
- Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. “Fitting

- Linear Mixed-Effects Models Using lme4.” *Journal of Statistical Software* 67 (1): 1–48. doi:10.18637/jss.v067.i01.
- Kondratyev, Alexander, and Elmira Zaynagutdinova. 2008. “Greater White-Fronted Geese (*Anser Albifrons*) and Bean Geese (a. *Fabalis*) on Kolguev Island—abundance, Habitat Distribution, and Breeding Biology.” *Vogelwelt* 129: 326–33.
- Kruckenberger, Helmut, Alexander Kondratyev, J MOOIJ, Christoph Zöckler, and E Zarigudinova. 2008. “White-Fronted Goose Flyway Population Status.” *Angew. Feldbiol* 2 (1): 77.
- Nolet, Bart A, Silke Bauer, Nicole Feige, Yakov I Kokorev, Igor Yu Popov, and Barwolt S Ebbinge. 2013. “Faltering Lemming Cycles Reduce Productivity and Population Size of a Migratory Arctic Goose Species.” *Journal of Animal Ecology* 82 (4). Wiley Online Library: 804–13.
- Wood, Simon N. 2003. “Thin-Plate Regression Splines.” *Journal of the Royal Statistical Society (B)* 65 (1): 95–114.
- . 2004. “Stable and Efficient Multiple Smoothing Parameter Estimation for Generalized Additive Models.” *Journal of the American Statistical Association* 99 (467): 673–86.
- . 2006. *Generalized Additive Models: An Introduction with R*. CRC press.
- . 2011. “Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models.” *Journal of the Royal Statistical Society (B)* 73 (1): 3–36.
- . 2013. *Generalized Additive Models: An Introduction with R*. 2nd ed. Chapman; Hall/CRC.