A hidden Semi-Markov model for inferring the structure of migratory bird flyway networks

Sam Nicol[1], Marie-Josée Cros[2], Nathalie Peyrard[2], Régis Sabbadin[2], Ronan Trépos[2],

[1] *CSIRO Land and Water, Dutton Park QLD, Australia*

[2] *INRAE, UR MIAT, F-31320 Castanet-Tolosan, France*

Every year, more than 50 million shorebirds migrate from overwintering habitat in Australasia to breeding grounds in Siberia. This migration is threatened by development pressures at stopover sites along the migratory route. Prioritising the conservation of critical bird habitat requires knowledge of the routes followed by birds but prioritisation remains difficult without knowing how sites are connected.

We propose to use statistical modelling and efficient inference tools to determine the most likely network of migration routes based on observed counts at stopovers. In practice, available count data are noisy due to irregular collection intervals and detection errors. We model the network, noisy data and duration of stopover at sites as a set of Hidden Semi Markov Models (HSMM) factored by the observations. The semi-Markov framework allows us flexibility to consider a sojourn time distribution other than the geometric distribution. This allows us to better fit to expert knowledge of sojourn times of birds. In the *FlywayNet* model, individual birds sojourn in stopover sites for a random period of time (according to a shifted Poisson distribution) before moving to other sites, according to an unknown multinomial distribution that we aim to estimate. Each observed count, at a given site and a given time, is assumed to follow a Poisson distribution with mean equal to the true number of birds. For this kind of HSMM, exact application of existing estimation methods based on the likelihood is not possible for even a small number of sites. It is due to the dimension of the hidden state and and to the dependencies between the hidden chains of each bird's trajectory conditionally to the observed counts. We designed two simulation-based methods to estimate the model's parameters, based on Monte Carlo EM [1] and Approximate Bayesian Computing (ABC, [2]) methods.

We compared the efficiency and quality of estimation of these algorithms on synthetic data simulated with the *FlywayNet* model with different numbers of sites, network structures, and numbers of missing observations. We observed very low errors for both algorithms: over the 300 problems simulated, the mean value of the mean absolute error over all parameters rescaled between 0 and 1 was of 0.08 for ABC and 0.06 for MCEM. As expected we observed an increase of the error when the number of sites or the maximal number of neighbors increased, but in all cases the mean value of the mean absolute error was always below 0.21. We observed no effect of the percentage of missing observations (we tested 30 % and 50 % of missings), probably because trajectories with missing observations can be easily interpolated.

Then we illustrated the behavior of the model and the estimation algorithms on applied case study using citizen science count data of the Far Eastern curlew (*Numenius madagascariensis*) in the East Asian-Australasian Flyway. Little is known about how Far Eastern curlews use stopover sites [4] or how the individual sightings data can be extrapolated to the population level. We modeled the flyway using a network of 8 sites representing the major known stopover regions for Far Eastern curlews. Count data at each site was determined by combining the information of the eBird[1] cheklists available in the site area, with post-treatment to reduce spatial bias due to difference in sampling effort. We applied the HSMM model on the eBird data from 2019. There are 32,000 birds with initial repartition derived from the observed count at time zero. As opposed to the experiments on simulated data, the estimators provided by ABC and MCEM were different. ABC results are more coherent with expert knowledge. For real data, it may be more robust than MCEM that deviates from the assumed *FlywayNet* model.

In conclusion, the *FlywayNet* model is an original model that is the first extension of Factorial HMM [3] to the hidden semi-Markov case and the first use for migratory network inference. Inference for such graphical models is complex but we designed two simulation-based algorithms which recovered parameters with good accuracy. The advantage of our approach is that it makes it possible to estimate flyway based only on limited count data at stopover sites. However, in a future work, it may be worth relaxing the following assumptions: the sojourn time is independent of the previous visited site and birds do not fly backward. One limitation of the MCEM and ABC algorithms could be their computational time. To circumvent this problem, we are currently investigating a VBEM algorithm, which should be faster with, hopefully, a quality of estimation maintained.

References

[1] C. Andrieu, N. De Freitas, A. Doucet, and M.L. Jordan. An introduction to MCMC for machine learning. *Machine Learning*, 50:5–43, 2003.

[2] Katalin Csilléry, Michael G. B. Blum, Oscar E. Gaggiotti, and Olivier François. Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology and Evolution*, 25(7):410–418, 2010.

[3] Z. Ghahramani and M. Jordan. Factorial hidden Markov models. *Mach. Learn.*, 29(2-3):245–273, 1997.

[4] Clive Minton, Rosalind Jessop, P Collings, and R Standen. The migration of eastern curlew numenius madagascariensis to and from australia. *Stilt*, 59:6–16, 2011.

[1] ebird.org