SEPARATING MORTALITY AND EMIGRATION: MODELLING SPACE USE, DISPERSAL AND SURVIVAL WITH ROBUST-DESIGN SPATIAL-CAPTURE-RECAPTURE DATA

Torbjørn Ergon (1)* and Beth Gardner (2)

(1): Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, P.O. Box 1066 Blindern, N-0316 Oslo, Norway

(2): Department of Forestry and Environmental Resources, North Carolina State University, Campus Box 7646, Raleigh, NC 27695-7646, USA

* Corresponding author: torbjorn.ergon@bio.uio.no

Appendix S2: Details of the Case study; Data, modelling procedure and posterior predictive goodness-of-fit tests

Data

To illustrate the potential use of our robust design spatial capture model, we chose a dataset on field voles (*Microtus agrestis*) from Kielder forest area in northern England (Ergon & Lambin 2013), and aimed to describe sex differences in space-use, survival and dispersal among adult, reproductively active individuals. For simplicity we only used data from a single summer season (year 2000) in one trapping grid. This trapping grid was located on a large, relatively homogenous, forest clear cut and had no "hard" habitat borders. Hence, the assumption that dispersal is equally likely in all directions and homogenous across space is not unreasonable. For a detailed description of the study system see Ergon *et al.* (2011) and references therein.

The trapping grid consisted of 192 traps in an 11 by 18 grid with traps at 7 meter intervals. Traps that can capture multiple individuals, Ugglan Special traps (Grahn AB; http://grahnab.se/en/fallor/), were baited with whole barley grains and carrots, and captured individuals were marked with unique Hauptner[TM] ear tags. Traps were checked at about 12 hour intervals at around 6 am and 6 pm such that each primary trapping period had respectively 3, 5, 4 and 5 secondary trap checks. We restricted the

analysis to adult individuals (118 females and 40 males) during the main reproductive season, with four primary trapping periods at intervals of 21 to 23 days between 10th of June and 15th of August.

Modelling procedure

Because we selected a homogenous group of individuals from only one season at one site, we assumed that all model parameters remained constant over time. Initial analysis showed that capture probability differed largely between morning and evening trapping sessions. Hence, we used the function for capture hazard rate given in eq. 4 of the main paper with an additive difference in $\ln(\lambda)$ between sexes and morning/evening trapping session. Since we aimed to describe sex differences, all other parameters were made sex-specific.

We fitted the model with six different dispersal models (parameters d in eq. 6 of the main paper): exponential, gamma and log-normal, with zero-inflated versions for each of these distributions. The zero-inflated models also included an offset parameter, being the minimum dispersal distance given dispersal. That is, the zero-inflated models for dispersal can be described as

$$d_{ik} = u_{ik}(d_{0,\text{sex}(i)} + \delta_{ik}),$$ 1.

where $u_{ik} \sim Bern(\psi_{\text{sex}(i),k})$ is the zero-inflation term, $d_{0,\text{sex}(i)}$ is the sex-specific offset parameter and δ_{ik} is either exponential, gamma or log-normally distributed with sex-specific parameters. This model is intuitively attractive because one should expect that individuals will often not move between primary sessions (e.g., if they have young in the nest), and given that they move, they should move a minimum distance. This model approaches the model without zero-inflation when $d_{0,\text{sex}(i)}$ is small and $\psi_{\text{sex}(i)}$ approaches 1. To avoid any such confounding, we set a lower bound of 10 meters in the prior distribution of $d_{0,\text{sex}(i)}$. In this way, the zero-inflation term $\psi_{\text{sex}(i)}$ can be interpreted as the probability of "non-negligible dispersal" between primary sessions.

Both survival probability and dispersal probability were scaled according to the length of the time-interval between trapping sessions as $\phi_{ik} = (\phi_{\text{sex}(i),k})^{t_{k+1}-t_k}$ and $\psi_{\text{sex}(i),k} = 1 - (1 - \psi_{\text{sex}(i)})^{t_{k+1}-t_k}$, respectively, where $t_{k+1} - t_k$ is the length of time between the start of primary session k and the start of primary session $k + 1$ (i.e., assuming that time-averaged mortality and dispersal hazard rates were the same in all intervals). Note, however, that time between primary sessions only varied from 21 to 23 days.

For comparison we also fitted a corresponding non-spatial robust design model with temporary emigration, as well as a standard CJS model, to the data. All models were fitted with Bayesian MCMC methods in the program JAGS (Plummer 2012) run through R by using the R-package 'rjags' (Plummer 2013). We used flat (uniform) priors within reasonable ranges for all parameters - see the JAGS model specification code in Appendix S2. The posterior for each model was sampled with at least 200,000 MCMC iterations following a burn-in of 10,000 iterations in 5 parallel runs. We confirmed chain convergence with the potential scale reduction factor ($\hat{R} < 1.01$; Brooks and Gelman (1998) and Gelman and Rubin (1992)).

Goodness-of-fit – Posterior predictive checks

We used simulation based posterior predictive checks (Gelman *et al.* 2004, chap. 6) to assess goodness-of-fit of the three components of the model. The capture component of the model was assessed by comparing the distributions of number of unique traps that each individual was captured in during each primary session with posterior predictive distributions of the same quantity. For assessment of the survival and dispersal components of the model, we compared observed and simulated posterior distributions of changes of mean capture locations between primary sessions as well as proportions of individuals known to be alive at each primary session that were ever recaptured. Simulations were based on the following procedure:

1. Draw a parameter vector with replacement from the posterior distribution (all parameters drawn from the same step in the MCMC simulations).
2. Simulate a dataset based on this parameter vector, using the exact same trap configuration and numbers of new individuals of each sex in each primary session as in the observed data. The first individual centres of activity were taken to be the mean location of the traps in which the individuals were captured during its first primary session.
3. Compute test statistics and p-values that compares specific aspects of the observed data with the simulated data.
4. Repeat steps 1 to 3 over 10,000 iterations and examine distribution of test statistics and p-values to assess goodness-of-fit.

This approach is somewhat different from standard posterior predictive checks (Gelman *et al.* 2004) in that we not only compare test statistics that are computed independently from the observed and the replicated data – we also compare the observed dataset with each of the simulated datasets with frequentist tests that compare specific distributions in the data that cannot be summarized by a single

test statistic (see below). If the observed data was independent from the data used in the model fitting, the expected distribution of p-values from these tests should be Uniform(0,1) under the null-hypotheses, but one should expect discrepancies from this when the same dataset is used for model fitting and model checking and when sample size is low. Nevertheless, a high mass at low p-values will indicate ways in which the model fits the observed data poorly.

We assessed the goodness-of-fit of each of the three model components as follows:

The spatial-capture-recapture component: For each primary session we tallied the number of unique traps in which each individual captured at least once in the primary session had been caught. We treated these counts as nominal data and tested for independence in a 2 by J contingency table (where J is the number of secondary sessions), using a Pearson Chi-square test (Zar 1984, chap. 9.12). We also combined the Pearson Chi-square statistics from all primary sessions in an overall test. The tallies from each of the simulated datasets were saved to enable subsequent investigation of any lack of fit.

The survival component: For each primary session we counted the number of individuals known to be alive and the proportion of these that were seen at any later primary session. Differences between the observed and simulated datasets in the probability of being seen at a later primary session was assessed by a binomial generalized linear model, testing for the interaction effect of 'set' (observed vs. simulated) and 'primary session' with a likelihood ratio test. The simulated distribution of the odds of being seen again for each primary session could subsequently be used to assess the survival component at each primary session by posterior predictive p-values (Gelman *et al.* 2004).

The dispersal component: For each primary session and individual, we computed the mean location of the traps in which the individual was captured, and used this to compute "face-value" dispersal distances from primary session k to primary session $k+1$ for those individuals that were captured in both session k and $k+1$ (in our data, recapture probability was very high, but when recapture probabilities are lower, one could instead calculate the distance moved before next capture). The distribution of these "face-value" dispersal distances in the observed data and each of the simulated datasets were compared with a two-sample Kolmogorov-Smirnov test for continuous data (function 'ks.test' in R (R Core Development Team 2013)). This test may also be applied separately to each primary session. We saved the quantiles (at 1% intervals) of each simulated distributions and used quantile-quantile plots to evaluate the nature of any lack of fit.

Each of these goodness-of-fit tests may reflect a number of model assumptions, and there may be aspects of the model that fits the observed data poorly that are not discovered by these tests. The purpose of the model checking is not to reject or accept the model ("all models are wrong" Box (1976)) but rather to understand the ways in which it does not fit the data.

References

Box, G.E.P. (1976) Science and statistics. *Journal of the American Statistical Association,* **71,** 791-799.

Brooks, S.P. & Gelman, A. (1998) General methods for monitoring convergence of iterative simulations. *J. Comp. Graph. Stat.,* **7,** 434-455.

Ergon, T., Ergon, R., Begon, M., Telfer, S. & Lambin, X. (2011) Delayed density-dependent onset of spring reproduction in a fluctuating population of field voles. *Oikos,* **120,** 934-940.

Ergon, T. & Lambin, X. (2013) Data from: Separating mortality and emigration: Modelling space use, dispersal and survival with robust-design spatial-capture-recapture data. Dryad Digital Repository. URL http://dx.doi.org/10.5061/dryad.r17n5.

Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B. (2004) *Bayesian Data Analysis,* 2nd edn. Chapman & Hall/CRC.

Gelman, A. & Rubin, D.B. (1992) Inference from iterative simulation using multiple sequences. *Stat. Sci.,* **7,** 457-511.

Plummer, M. (2012) JAGS Version 3.3.0 user manual. Sourceforge <http://sourceforge.net/>.

Plummer, M. (2013) Package 'rjags'. The comprehensive R archive network <http://cran.r-project.org/>.

Team, R.C.D. (2013) R version 3.0.1 The R Foundation for Statistical Computing.

Zar, J.H. (1984) *Biostatistical Analysis,* Second Edition edn. Prentice-Hall, Elgelwood Cliffs, New Jersey.