Ecography

E5236

Aarts, G., Mackenzie, M., McConnell, B., Fedak, M. and Matthiopoulos, J. 2008. Estimating space-use and habitat preference from wildlife telemetry data. – Ecography 31: 140–160.

Supplementary material

Appendix. Removing the geographical and environmental gridding

To derive a model that does not require grids in geographical or environmental space, we envisage two distinct, limiting operations: Operation 1 (O1) involves indefinitely increasing the resolution of the spatial grid. Operation 2 (O2) is an arbitrary increase in the resolution of the grid in environmental space. We first increase the resolution of the spatial grid (O1). Eventually, this will lead to most spatial grid cells being empty of observations and occupied cells containing, at most, one observation. Thus the number of telemetry observations occurring in the ith habitat tends to the number (u_i) of cells of that habitat that contain an observation. Using eq. S1, we can rewrite preference as

$$w_i = \frac{u_i}{a_i} = \frac{u_i}{\left(o_i + u_i\right)c} \tag{S1}$$

where o_i is the number of unoccupied spatial-grid cells of the ith habitat and c is the area of a cell in the spatial grid. Hence, while, at first, some cells may contain more than one observation (e.g. because individuals repeatedly use the same den or nest), as the resolution of the grid increases, even minute variations in position (such as those caused by observation error) will ensure that different observations are found in different cells. These variations will have no effect on the parameter estimates of the habitat preference model, because the environmental conditions in increasingly proximate cell-centrepoints will also be increasingly similar. Operation O1 leads to an ever-increasing number of cells to a level beyond practical use. To overcome this, we use a case-control approach (Prentice and Pyke 1979, Stephenson et al. 2006), originally devised for the analysis of rare diseases and designed to deal with sets of presence-only data (the cases). To obtain the control data, an arbitrary number of absences are retrospectively selected from the same population as the “cases”. In the context of telemetry studies, these are randomly selected points in space. Because O1 leads to an increasing number of unoccupied cells and a finite number of used cells, the probability of selecting a used cell tends to zero. The asymptotic theory on case-control (Prentice and Pyke 1979) ensures that, if the sample of controls is a) sufficiently large compared to the number of telemetry locations and b) representative of all accessible points in geographical space, the estimates of all coefficients (except the intercept) will not be sensitive to sample size.

We denote by k_a the total number of spatial-grid cells selected as controls and by p_i the probability of selecting a cell of the ith habitat. We assume that this probability is given by the relative availability of that habitat

$$p_i = \frac{a_i}{\sum_{\text{all } j} a_j} \qquad \text{(S2)}$$

Then, $k_a p_i$ gives the expected number of cells of habitat $\mathbf{i}$ contained in the control. The expected proportion of used cells of the ith habitat in the case-control sample is

$$h_i = \frac{u_i}{k_a p_i + u_i} \qquad \text{(S3)}$$

Comparing the values of these proportions for any two given habitats as k_a gets very large yields

$$\lim_{k_a \to \infty} \frac{h_1}{h_2} = \lim_{k_a \to \infty} \frac{(u_1 u_2)/k_a + u_1 p_2}{(u_1 u_2)/k_a + u_2 p_1} = \frac{u_1/a_1}{u_2/a_2} = \frac{w_1}{w_2} \qquad \text{(S4)}$$

indicating that, under the case-control paradigm, the quantity h_i defined in eq. S3 can be treated as proportional to preference,

$$\lim_{k_a \to \infty} h_i \propto w_i \qquad \text{(S5)}$$

We can model the observed number of presences in the case-control sample of cells from the ith habitat as a realization from a Binomial process with probability h_i and number of trials n_i (the total number of cells in the case-control data set that belong to the ith habitat).

$$\hat{u}_i \sim B(n_i, h_i) \tag{S6}$$

Now, consider performing O2. This will eventually lead to each cell in space being a unique habitat. Conversely, habitat i in environmental space is either present or absent in geographical space,

$$n_i \in \{0,1\} \tag{S7}$$

The implication of $n_i = 0$ is that this habitat was not available to the animal and is therefore not considered in further analysis. Thus, the process in eq. S6 becomes a Bernoulli

$$\begin{aligned} \hat{y}_i &\sim B(1, h_i) \\ h_i &= g^{-1}(\eta_i) = \frac{e^{\eta_i}}{1 + e^{\eta_i}} \\ \eta_i &= \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_j x_{i,j} \end{aligned} \tag{S8}$$