Multi-scale occupancy model vs. random effect parameterization

February 25, 2019

Prepared by Kathi Irvine & Wilson Wright

Follow-up re several discussions about occupancy models for the NABat sampling design.

NABat sampling design

Briefly, NABat proposes a "master sample" to facilitate collaboration and data-sharing among partners (Larsen et al. 2008). This approach allows different organizations to define their own spatial domain of interest and level of survey effort within a common probabilistic framework.

The NABat master sample was a probabilistic design created by applying the generalized random tessellation stratified (GRTS) algorithm to randomly order *all* sample units in the CONUS areal frame (Stevens and Olsen 2004, Loeb et al 2015). The NABat master sample was a spatially balanced ordered list of all the 10-km x 10-km grid cells and any consecutively numbered subset should be spatially balanced.

Within each grid cell i, the advice is to deploy 4 detectors in suitable habitat with one in each spatial sub-unit (5-km x 5-km quadrant). We denote the detector location or station with j. Then the detector records bat calls for one up to four consecutive nights (nightly survey denoted as k).

Multi-level occupancy models

p. 602-603 in Kery and Royle "Applied Hierarchical modeling in ecology" book describes two approaches: a top-down versus bottom-up analysis.

bottom-up

A bottom-up analysis fits the following model (i = grid cell, j = station or detector location within grid cell i, and k = a detector night at station j):

$$a_{ij} \sim \text{Bernoulli}(\theta_{ij}) \text{logit}(\theta_{ij}) = X_j\beta + \gamma_i y_{ijk}|a_{ij} = 1 \sim \text{Bernoulli}(p_{jk})\text{logit}(p_{jk}) = W_{jk}\alpha \tag{1}$$

where $\gamma_i \sim \text{Normal}(\mu, \sigma^2)$ is the grid-cell random effect that accounts for multiple stations within a grid cell, but is more a nuisance parameter than of ecological interest.

With this "bottom-up" model the ecological state of a grid cell (species occurrs or not), is driven by the local-station process. SO to map a prediction at a grid cell, we would have to derive the probability of occurrence based on the sub-units. IF we assume a station is a point-location, a prediction at a larger areal unit is tenuous (point process to areal unit). IF we assume $a_{ij} = 1$ is the occurrence state of a smaller areal sub-unit (e.g., 5-km x 5-km), then we could derive the state of the grid cell $Z_i = 1$ as the Pr(at least one $a_{ij} = 1$ for unit i). BUT this requires information at that level of sampling (5-km x 5-km) or values of X_j for all j sampled or not.

top-down

The top-down is the multi-state model proposed by Nichols et al. 2008.

A top-down analysis fits the following model (i = grid cell, j = station or detector location within grid cell i, and k = a detector night at station j):

$$Z_i \sim \text{Bernoulli}(\psi)a_{ij}|Z_i = 1 \sim \text{Bernoulli}(\theta_{ij})y_{ijk}|a_{ij} = 1 \sim \text{Bernoulli}(p_{jk}) \tag{2}$$

You can think of this as "top-down" because the process at governing Z_i is defining the local states at a station a_{ij}. Currently, when we are fitting models we are unable to estimate θ separate from p because we

don't have a lot of information to explain the heterogeniety in θ instead we are estimating $p' = \theta * p$. IF the design has four spatial replicates without temporal replication we can't estimate separately. BUT if we retain both stations and nights, we could (I think).

One strong appeal about trying to use the multi-state model is that it maintains the connection to spatial predictions at the grid cell level and allows for informing local-availability or occurrence. This model could very well solve the issue of motivating NABat users to contribute if we can estimate their local processs ($a - level$) and their data can contribute to a larger regional predictive map.