

## Estimating percent-area-occupied and related dynamics from presence-absence data.

Estimation of occupancy rates and associated dynamics (i.e. extinction and colonization) from presence-absence data is fundamental to many habitat models (Cabeza *et al.* 2004), metapopulation studies (Hanski & Gilpin 1997) and monitoring efforts. Such studies are often interested in whether a defined ‘patch’ of ‘site’ is occupied by one or more individuals of a population or species. In general data are collected by visiting a number of sites and recording whether any individuals of interest are present (recorded as a 1) or not (recorded as a 0). Occupancy probability is often calculated as the proportion of sites that are occupied; extinction probability as the proportion of occupied sites at time  $t$  not occupied at time  $t+1$ ; and colonization probability as the proportion of sites not occupied at time  $t$  occupied at time  $t+1$ . However such presence-absence data, and resulting estimates, can be confounded by detection error; namely that a recorded ‘absence’ may in fact be a ‘non-detection’ of available individuals, and not a true absence. Such data should more precisely be referred to as presence/apparent absence data and not presence-absence data. Using such data with naïve estimators will most likely result in under-estimates of occupancy and colonization probabilities and over-estimates of extinction probabilities. If a detection probability can be calculated, then unbiased estimators of occupancy, extinction and colonization can be derived. MacKenzie *et al.* (2002; 2003) are the seminal papers describing such estimation methods. The recent book, MacKenzie *et al.* (2006) details these papers as well as other developments and is a must read for anyone interested in these methods. There is a good chapter on study design. Practically, these occupancy methods share much in common with the closed population abundance estimators and the ‘robust design’ (Pollock *et al.* 1990). Below we will focus on the estimation of occupancy probabilities.

Underlying assumptions in estimating occupancy are 1) surveyed sites are occupied by the species of interest throughout the duration of the study, with no sites becoming occupied or unoccupied during the survey period (i.e., the system is closed), 2) species are not falsely detected, but can remain undetected if present, 3) species detection at a site is assumed to be independent of species detection at other sites. For example a number of ponds may be visited to assess occupancy rates of salamanders. The surveys take place during a short period during the breeding season when ponds (sites) can assume to be occupied. Each pond is visited once a day for five days (or alternatively five locations within a pond could be sampled) and whether any salamanders are present at each pond is recorded. Data for a single pond ( $i$ ) can take the form of an ‘encounter history’ such as:

01010

Where ‘0’ signifies an apparent absence and a ‘1’ signifies a detection. Since salamanders were detected at least once, we assume the site was occupied across all five sampling occasions, but not detected on sampling occasions 1, 3 and 5. If we denote  $\psi$  as occupancy probability and  $p$  as detection probability we can designate the likelihood for the above encounter history as:

$$\psi_i(1 - p_1)p_2(1 - p_3)p_4(1 - p_5).$$

Not detecting a species does not imply absence. An encounter history of:

00000

could occur even if a site is occupied. This is an important difference when compared to the classical abundance estimation models in which such capture histories are never observed. In the occupancy arena, the site is the observation unit, not the individual animal. The likelihood statement for this capture history would be:

$$\psi_i(1 - p_1)(1 - p_2)(1 - p_3)(1 - p_4)(1 - p_5) + (1 - \psi_i)$$

read as the site was either occupied by salamanders but these salamanders were not detected over the five sampling occasions (the first term) or the site was not occupied (the second term). We can write this multinomial more efficiently as:

$$\psi_i \prod_{k=1}^5 (1 - p_k) + (1 - \psi_i).$$

The model can be extended to include covariates using a logistic model as we have done with other mark-recapture models in the workshop. Note that because occupancy does not change over time, appropriate covariates for occupancy would be time constant and site specific while covariates for detection could be time varying. This is in contrast to classic mark-recapture models where individual time-varying covariates can not be used (remember in the classic case individuals are not observed on all occasions and thus time-varying covariates would not be available).

In Program Mark missing data are also easily handled. In the above example, if the pond was not visited on the third day the encounter history would be:

01.10

with the ‘period’ denoting no data collected. The likelihood would be:

$$\psi_i(1 - p_1)p_2p_4(1 - p_5)$$

indicating no information is available to estimate  $p_3$ .

### **Further extensions and readings**

The occupancy model has been extended to estimate extinction and colonization over multiple years using a ‘robust design’ approach (MacKenzie *et al.* 2003). If our salamander pond was visited three years in a row, the encounter history may be:

01010 000.0 10111

with the three sets of five digits representing the three years. Occupancy rates for each year are calculated as well as extinction ( $\epsilon$ , Greek epsilon) and colonization ( $\gamma$ , Greek gamma) between years. In year two no salamanders were detected (note the pond was not visited during the 4<sup>th</sup> sampling occasion of that year). This could represent an extinction event between year 1 and year 2, or it could represent a series of non-detections during year 2. The multinomial for this capture history can be tediously written out. Interested readers should consult MacKenzie *et al.* 2003 or MacKenzie *et al.* 2006). The advantage of estimating the extinction and colonization rates for sites is that the temporal dynamics of patch occupancy are available, providing more information than just the occupancy rate through time.

The occupancy model has also been extended to investigate whether the occupancy of one species influences the occupancy of a second species (i.e. co-occurrence). This development is detailed in (MacKenzie *et al.* 2004).

The paper by Royle and Nichols (2003) investigate the relationship between ‘patch’ or ‘site’ level detection probabilities and individual detection probabilities. If  $p_s$  represent the site-level detection probability and  $p_i$  represents an individual detection probability then the probability that a species was not detected at a site ( $1 - p_s$ ) is equal to the probability that an individual was not detected raised to the number of individuals ( $n$ ) at the site  $(1 - p_i)^n$  or:

$$1 - p_s = (1 - p_i)^n .$$

If a site-level detection probability can be obtained and an individual detection probability can be model, a latent estimate of abundance can be obtained from detection/apparent absence data.

This methodology continues to expand rapidly. More recent advances focus on incorporating false positives into the estimation (Royle and Link 2006), dealing with heterogeneity in detection probabilities (Royle 2006) and applying these methods to estimating species richness (MacKenzie *et al.* 2006, Dorazio *et al.* 2006).

### Literature Cited

- Cabeza M., Araujo M.B., Wilson R.J., Thomas C.D., Cowley M.J.R. & Moilanen A. (2004) Combining probabilities of occurrence with spatial reserve design. *J Appl Ecology*, 41, 252-262
- Dorazio, R. M., J. A. Royle, B. Soderstrom, and A. Glimskar. 2006. Estimating species richness and accumulation by modeling species occurrence and detectability. *Ecology* 87:842-854.

- Hanski I.A. & Gilpin M.E. (1997) *Metapopulation biology - ecology, genetics and evolution*. Academic Press, San Diego.
- MacKenzie D.I., Bailey L.L. & Nichols J.D. (2004) Investigating species co-occurrence patterns when species are detected imperfectly. *J Anim Ecology*, 73, 546-555
- MacKenzie D.I., Nichols J.D., Hines J.E., Knutson M.G. & Franklin A.B. (2003) Estimating site occupancy, colonization and local extinction when a species is detected imperfectly. *Ecology*, 84, 2200-2207
- MacKenzie D.I., Nichols J.D., Lachman G.B., Droege S., Royle J.A. & Langtimm C.A. (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83, 2248-2255
- MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. H. Pollock, L. L. Bailey, and J. E. Hines. 2006. Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence. Elsevier, Amsterdam.
- Pollock K.H., Nichols J.D., Brownie C. & Hines J.E. (1990) Statistical-Inference for Capture-Recapture Experiments. *Wildlife Monographs*, 1-97
- Royle, J. A. 2006. Site occupancy models with heterogeneous detection probabilities. *Biometrics* 62:97-102.
- Royle J.A. & Nichols J.D. (2003) Estimating abundance from repeated presence-absence data or point counts. *Ecology*, 84, 777-790
- Royle, J. A., and W. A. Link. 2006. Generalized site occupancy models allowing for false positive and false negative errors. *Ecology* 87:835-841.