

Community Occupancy Model

Brian D. Gerber

November 12, 2024

Species Richness and Joint Modeling while Accounting for Detection Probability

The core idea of a community occupancy model is to jointly model detection data (Y) for species s (total n species observed) at site i with replicate surveys J to estimate the total species richness and share information across species in terms of effects on occurrence of each species as,

$$\begin{aligned} Y_{s,i} &\sim \text{Binomial}(p_{\mu_{s,i}}, J) \\ p_{\mu_{s,i}} &= p_i \times Z_{s,i} \\ Z_{s,i} &\sim \text{Bern}(\psi_{\mu_i}) \\ \psi_{\mu_i} &= \psi_i \times w_i \\ w_i &\sim \text{Bern}(\omega). \end{aligned}$$

The parameter ω is an inclusion probability of species that have not been detected. This parameter governs the latent parameter w_i to decide on whether a species should be included as part of the community or not. Species that are detected are included and species that have not been detected across any sites are included with probability ω . For this to work, the detection data Y needs to include the sampled detection data of observed species along with all zero detections for n_0 number of species that have not been detected but could have been part of the community; this statistical trick is known as parameter-expanded data augmentation and is quite common for Bayesian models that incorporate detection probability (Royle and Dorazio, 2010). *A priori*, there should be a known upper limit to the number of species in an area that could be possibly detected as part of the community taxa that can be sampled with the given methodology. For well known species (e.g., vertebrates), this can be decided on based on knowledge of the area and species distribution maps and knowledge of the sampling methodology. However, statistically, you could include any really large number that would lead to estimating species richness beyond realistic values of species richness of the area.

The occupancy and detection parameters for each species (ψ_s , p_s , respectively) can still be modeled with a logit-linear model, as is done with single species occupancy models.

The total species unobserved is summarized as,

$$\sum w_{n_0:(n+n_0)}.$$

Reading

- For original formulations, see Dorazio et al. 2006 and 2010.
- For a review of these types of models, see Devarajan et al. 2020.
- For a focus specifically on estimating species richness with this type of model, see Tingley et al. 2020.

References

- Devarajan, K., Morelli, T.L. and Tenan, S. (2020), Multi-species occupancy models: review, roadmap, and recommendations. *Ecography*, 43: 1612-1624. <https://doi.org/10.1111/ecog.04957>
- Dorazio, R. M., Kery, M., Royle, J. A., & Plattner, M. (2010). Models for inference in dynamic metacommunity systems. *Ecology*, 91(8), 2466-2475.
- Dorazio, R. M., Royle, J. A., Söderström, B., & Glimskär, A. (2006). Estimating species richness and accumulation by modeling species occurrence and detectability. *Ecology*, 87(4), 842-854.
- Royle, J. A., & Dorazio, R. M. (2012). Parameter-expanded data augmentation for Bayesian analysis of capture–recapture models. *Journal of Ornithology*, 152(Suppl 2), 521-537. <https://link.springer.com/article/10.1007/s10336-010-0619-4>
- Tingley, M. W., Nadeau, C. P., & Sandor, M. E. (2020). Multi-species occupancy models as robust estimators of community richness. *Methods in Ecology and Evolution*, 11(5), 633-642.