Proceedings of the 11th International Coral Reef Symposium, Ft. Lauderdale, Florida, 7-11 July 2008
Session number 18

Occupancy Models for Estimating the Size of Reef Fish Communities

M.A. MacNeil,[1,4] C.J. Fonnesbeck,[2] and T.R. McClanahan[3]

1) National Research Council, NOAA National Marine Fisheries Service, Panama City Laboratory, 3500 Delwood Beach Rd., Panama City Beach, FL 32408, USA
2) Department of Mathematics and Statistics, University of Otago P.O. Box 56, Dunedin, New Zealand
3) Marine Programs, Wildlife Conservation Society, 2300 Southern Blvd., Bronx, NY 10460, USA
4) Australian Institute of Marine Science, PMB 3, Townsville MC, 4810 Queensland, Australia

Abstract. Estimating reef fish species richness from underwater visual census (UVC) data has the potential to be negatively biased by non-detections of species present at a given site when the probability of detecting a species is less than one. To mitigate the effects of such sampling biases on subsequent ecological conclusions, we develop an occupancy model framework that estimates the probabilities of both detection and presence at locations in coastal Kenya using covariate logit-linear models. Based on a capture-mark-recapture (CMR) data structure for repeat observations over a closed sampling period, we find that reef-fish species characteristics are informative about differences in observed detection probabilities. Specifically, we show that schooling behaviour leads to higher probabilities of detection; that larger fishes are, on average, less detectable; and that detectability varies greatly by functional group. We also show how these characteristics of detection were poorly correlated with probabilities of occupancy and find that site-based covariates such as protection status may provide better models for presence. Finally we discuss the use of occupancy models in reef-fish surveys and suggest their use may be an important tool for reef ecologists in the future.

Key Words: mark-recapture; diversity; species richness; underwater visual census

Introduction

The diversity of coral reef fishes is unique in the marine environment and is among the characteristics that will most rapidly reflect fundamental changes in the ecological function of reef ecosystems. Although many alternate indices of diversity have been devised and used successfully in ecology (Buckland et al. 2005), species richness remains a prominent metric to describe the state of a given community (Kere and Schmid 2006). While species richness is frequently used to underpin the development of ecological theory, observing the true species richness of a given community is frequently difficult in practice (Dorazio et al. 2006) and, as a consequence of sampling error, the process signals contained in a given set of data may be weak or biased to an unknown degree (Allen and Starr 1982).

The potential for bias is particularly apparent in reef fish surveys where, for instance, small cryptic or large mobile species may not be readily observed due to low probabilities of detection in underwater visual census (UVC) data (Kulbicki 1998). Many reef ecologists recognize this issue and as a result the smallest and largest species on a given reef are frequently omitted from target species lists (McClanahan 1994). For the species that are present on target lists, detection probabilities are likely to be less than one (MacNeil et al. 2008b,a) and therefore fish species will have some probability of not being detected in a given survey, even over replicate transects at the same location. If detection probabilities are low, there may be many species present at a given location that remain undetected, negatively biasing the species richness observed to a substantial degree. Even though this issue is recognized conceptually, detectability is most often ignored during analysis and an implicit assumption is made that the probability of detection equals one.

Although most conventional sampling programmes do not address detectability directly in their design, model frameworks exist to address the issue of incomplete detectability during analysis. These models of species richness are based on capture-mark-recapture (CMR) models of tagged animals, whereby the probability of observing any given species on a particular transect is quantified relative to the detection histories of species observed during the survey, using a set of relevant covariates (Boulinier et al. 1998). Such an approach has been used successfully to estimate species richness in surveys of both birds (Kere and

Schmid 2006) and reef fish (MacNeil et al. 2008 a,b).

Although detection probabilities are important estimands in the quantification of species richness in UVC data, on their own they carry the implicit assumption that every species in the analysis was present at the sampling site during the census. For many species, this may not be the case, and the zeroes that arise in a set of UVC data can be due either to low probabilities of detection or because a given species was not present at a given sampling location. This second source of zeros has been called the occupancy state which, as it is either one (species present) or zero (species absent), can be quantified using presence-absence models combined with models for heterogeneous detection (MacKenzie et al. 2006). By combining models for detectability and occupancy, an analytical framework can be implemented to estimate species richness at a given survey location and, in the process, help to improve the ecological signal present in UVC data.

In this paper we use a well-known set of UVC data to demonstrate the conceptual utility of occupancy models to reef fish data. The statistical model presented describes separate components for handling the detectability and occupancy states of individual reef fish species, where covariate information from the species observed are used to estimate the proportion of a fixed species list that were not observed. The models are presented using a Bayesian inference framework.

Methods

To illustrate the applicability of an occupancy framework to reef fish UVC observations we selected data collected from coastal Kenya in the Western Indian Ocean (McClanahan et al. 2007). The collection of these data have been described elsewhere (McClanahan 1994) but briefly, the sampling scheme consisted of between 4 to 9 replicate UVC belt transects (5 m x100 m) conducted at four locations, two of which were protected from fishing (Mombasa; Malindi) and two of which were unprotected (Diani, Kanamai). Observations were made between 1992 and 2005, although sites were not surveyed in every year. The data were collected using discrete group sampling (DGS) where, depending on their physical and behavioural characteristics, fish from related groups or families were sampled on separate passes of a repeat transect in a given area. For these data, fish from eight families (Acanthuridae, Scaridae, Pomacentridae, Chaetodontidae, Pomacanthidae, Balistidae, Diodontidae, and Labridae) were counted to the species level using a fixed list of 161 potential species. Because the probability of losing or gaining new fish species among transects at a given site is near zero, the data permitted the mild assumption that the community was closed to species immigration and emigration over the sampling interval.

Occupancy models were built up from a conditional occupancy framework (MacKenzie et al. 2006) where, following conventional notation, y_i is the observed number of species detections accross k transects at a given site, ψ_i represents the probability of species i occupancy at a specific site, and θ_i represents the probability of detection for species j given that it is present at a particular site. The approach relates the occupancy state of species i to detectability using a two-part conditional model. First, if a species has not been observed ($y_i = 0$) then its occupancy state is unknown and can be described by

the sum of two conditional probabilities, the species being either not present ($\psi_i = 0$) or present ($\psi_i = 1$) but unobserved ($\theta_i < 1$):

$$p(y_i = 0) = \psi_i(1-\theta_i)^k + (1-\psi_i). \qquad (1)$$

Conversely, if the species is observed at least once over k transects ($y_i > 0$) then the record of events can be described as the bernoulli-distributed probability of occupancy times the binomial distribution of detections:

$$p(y_i > 0) = \psi_i \binom{k}{y_i} \theta_i^{y_i} (1-\theta_i)^{k-y_i}, \qquad (2)$$

for $y_i = 1, 2, ..., k$. Using these models as a basic structure for the detection of individual fish species, the framework can be extended to accommodate covariates of interest for ψ and θ using linear models and a logit link. For occupancy this is given by:

$$\frac{\psi_i}{(1-\psi_i)} = \beta_0 + \beta_1 x_1 + ... + \beta_j x_j, \qquad (3)$$

where the β's are covariate parameters for presence and similarly for detectabitliy:

$$\frac{\theta_i}{(1-\theta_i)} = \gamma_0 + \gamma_1 x_1 + ... + \gamma_j x_j, \qquad (4)$$

where the γ's are covariate parameters for detection. All of the γ's and β's are assumed to be $N(0,\sigma_j)$ distributed. Further details on this model and its implementation can be found in MacKenzie et al. (2006). Previous work on reef fish detectability has shown that schooling behaviour (schooling/non-schooling; SC); reef fish functional

group (FG; a grouping factor); and maximum total length (TL_{max}; a continuous variable) can affect probability of detection (MacNeil et al. 2008b,a). For simplicity, we used these three covariates in (4), the detection portion of our approach, to model the Kenyan reef-fish data and did not pursue covariate selection. In addition, we chose to include the same covariates in (3), as occupancy covariate selection has not been demonstrated previously and because our objective was simply to outline a form for the occupancy approach.

The resulting joint detection-occupancy model was implemented by Markov-chain Monte-Carlo (MCMC) simulation in the pymc (Fonnesbeck et al. 2008) toolkit for the Python programming language (Python Software Foundation 2006). Models were run for 15,000 iterations after a 15,000 iteration burn-in period and model convergence was assessed through visual inspection of parameter chains, where chains showing adequate mixing and stability were considered to have reached convergence.

Results

A range of between 34 and 84 species were observed across all transects, with consistently higher richness in the closed areas of Mombasa and Malindi. Observed occupancy proportions (Ω) were 23% or less in the open areas and were 34% or greater between the closed areas (Table 1). There was considerable year-to-year variation in richness values among sites, particularly for Mombasa which went from a high of 84 observed species in 1991 to a low of 55 observed species the following year.

Occupancy model convergence and mixing were good for all estimated parameters and all chains appeared to have reached convergence after the burn-in period. Posterior richness estimates ranged from 41% (66 spp.) to 74% (118 spp.) of the total species list, with estimates from 19% (31 spp.) to 41% higher than the observed proportions. There was slightly larger average estimated increases in richness in the closed (24%) versus the open (22%) areas. No sites showed evidence for all species being present, and the highest 95% credible interval bound lying at 83%.

Detection parameter estimates were highly variable among sites, with marked differences both between and among closed and open areas. Scrapevator detectability for example was relatively higher than site-averaged detectability in the closed areas, while being in the open areas. For other groups however (e.g. detritivores) detectability estimates were not consistent between management areas. The most detectable fishes tended to be grazers in the closed areas, and detritivores and invertivores in the open areas. Despite such evidence for differences among many functional groups, in many instances estimates spanned zero, suggesting additional covariates may be required to provide a satisfactory model of detectability at these sites.

Overall, the distribution of median estimated detection probabilities were more variable within open, as opposed to closed, areas. Kanamai in particular had a wider range of median detection probabilities across years relative to the other sites, with a somewhat diffuse distribution between zero and one that was apparent also in the Diani estimates. Conversely, the closed areas had more normally-distributed detection estimates, with median detectability values near 0.6 across years.

Unlike many of the detectability parameter estimates, fish characteristic parameter estimates were all poorly correlated with occupancy status, as almost all parameter estimates spanned zero. Exceptions to this included consistently lower planktivore occupancy in the open sites, and lower invertivore occupancy probabilities in Diani across all years. Although the covariate models were generally uninformative, the distribution of median estimated occupancy and detectability were consistently lower in the open versus closed areas. With the exception of Diani in 2006, the distribution of median estimated occupancy within sites was relatively consistent among years.

Median posterior estimates of detection and occupancy revealed distinct patterns in relation to median fish abundance per transect (Fig. 1). While sites with higher median abundance had clearly higher median probabilities of detection (Fig. 1a), median occupancy appeared (again with the exception of Diani in 2006) to be correlated with protection status rather than median transect abundance (Fig. 1b). To our knowledge, this has not been demonstrated previously.

Although detectability differences observed between open and closed areas are known to be highly correlated with abundance at a given sampling site (MacKenzie et al. 2006), probability of occupancy should be less so as a portion of this bias has already been expressed through the detection portion of the model. This appeared to be the case in our results, where only detection probabilities were shown to be highly correlated with median abundance at a given site.

Occupancy estimates on the other hand showed no such trend, instead reflecting unmodeled processes inherent in protection status that appear to have affected true species richness. Such processes may include effects of fishing on habitat quality that can reduce the occupancy probability of, for instance, coral-dependent species (McClanahan 1994). Indeed the lowest occupancy estimates consistently observed were for planktivorous species that, because they utilize the structural

complexity of the corals for shelter, are tightly linked to habitat conditions (MacNeil 2008) and likely to be directly affected by losses of reef habitat (Graham et al. 2008). The fine-scale resolution of species characteristics and their effects on detection and occupancy remain to be seen however, and we hope that the results presented here will generate new hypotheses in this area.

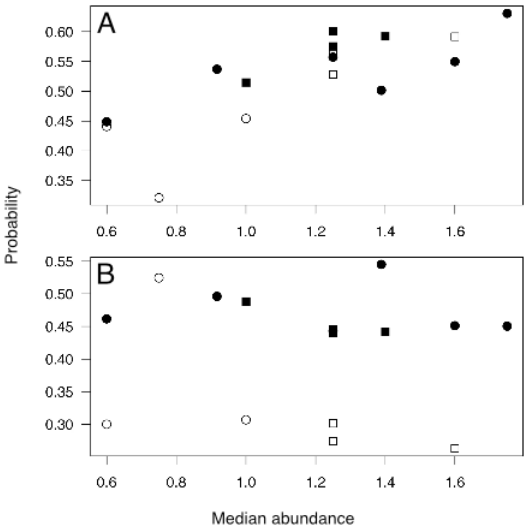

Figure 1: Median posterior detection (A) and occupancy (B) probabilities per median abundance of fishes observed per transect at Diani (open circle), Kanamai (open square), Malindi (filled square), and Mombasa (filled circle).

Given these results, we believe that occupancy estimates more directly reflect true trends in species richness through time than raw counts, as bias driven by the local abundance of fishes has, in part, been handled through modeling of detectability. Although these characteristics are evident in these data, it will require additional validation through both simulation and field studies to determine the efectiveness of occupancy models for informing management.

Discussion

We have outlined a general framework for estimating reef-fish species richness from conventional UVC data in which the issue of incomplete detection (a form of sampling error) is addressed directly using a conditional statistical model. The strength of the analysis is in the explicit parameterization of occupancy, where probabilities of presence can readily be linked to any potential covariates of interest on a given reef. While the occupancy models we developed from fish physical characteristics showed little correlation with occupancy probabilities, there is exceptional potential for occupancy models to link probabilities of presence with more likely covariates such as habitat characteristics (MacNeil 2008).

The covariates we chose, developed to deal with reef-fish detectability (MacNeil et al. 2008b,a) performed poorly in estimating occupancy, suggesting that fish physical characteristics are much more likely to affect detection probabilities than occupancy. If occupancy models are to be used routinely in reef-fish UVC surveys, covaraites must be identified that correlate appropriately with occupancy status. The presences or absences of particular reef fish species have been repeatedly linked to the site-scale habitats in which they are observed (Caley et al. 2001; Gratwicke and Speight 2005) and this suggests that reef-fish occupancy will most likely reflect trends in site-level characteristics. Although such an analysis has yet to be put into practice, our results show tentatively that protection status is a potentially informative covaraite, supporting the role of MPA's in promoting diversity.

That observed species richness is greatly affected by sampling bias due to the characteristics of the fishes observed has been widely recognized in reef fish surveys (Samoilys and Carlos 2000). Typically this is a downward bias (Ackerman and Bellwood 2000) that can, particularly where rare species are of interest, severely affect the conclusions drawn in a given analysis. In this example for instance, the highest and lowest observed species richness' in Mombasa were observed in subsequent years (1991-1992). Yet it is highly unlikely that there was a true loss of 29 species between those years followed by an increase of 9 species the year following. It is much more likely that low detectability of some of the rarer or more mobile species affected the observed counts across the transects within each year, leading to depressed counts in 1992. In contrast, the occupancy model estimates overlap substantially, indicating that there were likely between 113-121 species present at Mombasa in the early 1990's.

Given the consistent increases in observed species richness across all sites and years, using occupancy models to estimate true richness has substantial potential to reduce sampling error in a given set of data and, as a consequence, to increase the process signal strength that is inherent in the data itself (Allen and Starr 1982). Where, for example, the kinds of year to year variation discussed above are introduced to data through detection heterogeneity among species, the process signal of interest may become substantially weakened. Given that the most important questions for reef ecologists are ecological ones, these kinds of statistical estimators have the potential to structure some portion of sampling error in data,

while also providing ecological quantities of interest such as occupancy. These twin properties of reducing sampling variation and increasing signal strength suggest that occupancy models may see wide application in reef ecology in the near future.

Acknowledgements

This research was supported by funding from the National Research Council (USA) and the Wildlife Conservation Society. Permission to work in the marine parks was granted by the Kenya Wildlife Service and thanks goes to them for their many years of assistance.

References

Ackerman J, Bellwood D (2000) Reef fish assemblages: a re-evaluation using enclosed rotenone stations. Mar Ecol Prog Ser 206:227–237

Allen T, Starr T (1982) Hierarchy: Perspectives for Ecological Complexity. University of Chicago Press, Chicago, USA

Boulinier T, Nichols J, Sauer J, Hines J, Pollock K (1998) Estimating species richness: the importance of heterogenity in species detectability. Ecology 79:1018–1028

Buckland S, Magurran A, Green R, Fewster R (2005) Monitoring change in biodiversity through composite indicies. Phil Trans Roy Soc Lon B 360:243–254

Caley M, Buckley K, Jones G (2001) Separating ecological effects of habitat fragmentation, degredation, and loss on coral commensals. Ecology 82:3435–3448

Fonnesbeck CJ, Huard D, Patil A (2008) PyMC: Markov chain Monte Carlo for Python, version 2.0. http://code.google.com/p/pymc/

Graham NAJ, McClanahan TR, MacNeil MA, Wilson SK, Polunin NVC, Jennings S, Chabanet P, Clarke S, Spalding M, Letourneur Y, Bigot L, Galzin R, Ohman M, Garpe K, Edwards A, Sheppard C (2008) Climate warming, marine protected areas and the ocean-scale integrity of coral reef ecosystems. PLoS ONE, 3:e3039 doi:10.1371/journal.pone.0003039

Gratwicke B, Speight M (2005) The relationship between fish species richness, abundance and habitat complexity in a range of shallow tropical marine habitats. J Fish Biol 66:650–667

Kere M, Schmid H (2006) Estimating species richness: calibrating a large avian monitoring programme. J App Ecol 43:101–110

Kulbicki M (1998) How the aquired behaviour of commercial reef fishes may influence the results obtained from visual censuses. J Exp Mar Biol Ecol 222:11–30

MacKenzie D, Nichols J, Royle J, Pollock K, Hines J, Bailey L (2006) Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence. Elsevier, San Diego, USA

MacNeil MA (2008) Making empirical progress in observational ecology. Env Cons 25:1–4

MacNeil MA, Graham NAJ, Conroy MJ, Fonnesbeck CJ, Polunin NVC, Rushton SP, Chabanet P, McClanahan T (2008a) Detection heterogeniety in underwater visual census data. J Fish Biol *IN PRESS*.

MacNeil MA, Tyler E, Fonnesbeck CJ, Rushton SP, Polunin NVC, Conroy MJ (2008b) Acounting for detectability in reef-fish biodiversity estimates. Mar Ecol Prog Ser 367:249–260

McClanahan TR (1994) Kenyan coral reef lagoon fish: effects of fishing, substrate complexity, and sea urchins. Coral Reefs 13:231–241

McClanahan TR, Graham NAJ, Calnan J, MacNeil MA (2007) Toward pristing bioimass: reef fish recovery in coral reef marine protected areas in Kenya. Ecol Appl 17:1055–1067

Python Software Foundation (2006). The python programming language. http://python.org

Samoilys M, Carlos G (2000) Determining methods of underwater visual census for estimating the abundance of coral reef fishes. Env Biol Fish 57:289–304

Table 1: Model estimates by site for observed (N) and estimated (N_{hat}) reef fish species richness from coastal Kenya; Ω and Ω_{hat} indicate observed and estimated proportion of total species list (161 spp.) at each location. Estimates are median values (± 95% credible intervals); status (open/closed) refers to fishing.

Site	Year	Status	N	Ω	N_{hat}	Ω_{hat}
Diani	1992	open	34	0.21	71(57, 87)	0.44(0.35, 0.54)
Diani	2003	open	34	0.21	72(55, 96)	0.45(0.34, 0.60)
Diani	2006	open	37	0.23	103(81, 120)	0.64(0.50, 0.75)
Kanamai	1992	open	35	0.22	66(53, 81)	0.41(0.33, 0.50)
Kanamai	2003	open	33	0.21	66(52, 90)	0.41(0.32, 0.56)
Kanamai	2006	open	37	0.23	73(58, 91)	0.46(0.36, 0.57)
Mombasa	1991	closed	84	0.53	123(113,133)	0.64(0.70, 0.83)
Mombasa	1992	closed	55	0.34	101(86, 121)	0.63(0.53, 0.75)
Mombasa	1993	closed	64	0.40	104(92, 117)	0.65(0.57, 0.73)
Mombasa	1998	closed	63	0.39	105(93, 117)	0.66(0.58, 0.73)
Mombasa	2003	closed	79	0.49	118(108, 129)	0.74(0.67, 0.80)
Mombasa	2006	closed	67	0.42	108(96, 119)	0.68(0.60, 0.74)
Malindi	1992	closed	67	0.42	108(96, 120)	0.68(0.60, 0.75)
Malindi	1996	closed	70	0.44	110(99, 122)	0.69(0.61, 0.76)
Malindi	2003	closed	70	0.44	113(102, 126)	0.71(0.63, 0.78)
Malindi	2006	closed	62	0.39	105(93, 119)	0.66(0.58, 0.74)