

Bayesian Modelling
of Direct and Indirect Effects of
Marine Reserves on Fishes

A thesis presented in partial fulfilment
of the requirements for the degree of

Doctor of Philosophy

in

Statistics

at Massey University, Albany, New Zealand.

Adam Nicholas Howard Smith

2016

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Abstract

This thesis reviews and develops modern advanced statistical methodology for sampling and modelling count data from marine ecological studies, with specific applications to quantifying potential direct and indirect effects of marine reserves on fishes in north eastern New Zealand. Counts of snapper (*Pagrus auratus*: Sparidae) from baited underwater video surveys from an unbalanced, multi-year, hierarchical sampling programme were analysed using a Bayesian Generalised Linear Mixed Model (GLMM) approach, which allowed the integer counts to be explicitly modelled while incorporating multiple fixed and random effects. Overdispersion was modelled using a zero-inflated negative-binomial error distribution. A parsimonious method for zero inflation was developed, where the mean of the count distribution is explicitly linked to the probability of an excess zero. Comparisons of variance components identified marine reserve status as the greatest source of variation in counts of snapper above the legal size limit. Relative densities inside reserves were, on average, 13-times greater than outside reserves.

Small benthic reef fishes inside and outside the same three reserves were surveyed to evaluate evidence for potential indirect effects of marine reserves *via* restored populations of fishery-targeted predators such as snapper. Sites for sampling were obtained randomly from populations of interest using spatial data and geo-referencing tools in R—a rarely used approach that is recommended here more generally to improve field-based ecological surveys. Resultant multispecies count data were analysed with multivariate GLMMs implemented in the R package *MCMCglmm*, based on a multivariate Poisson lognormal error distribution. Posterior distributions for hypothesised effects of interest were calculated directly for each species. While reserves did not appear to affect densities of small fishes, reserve-habitat interactions indicated that some endemic species of triplefin (Tripterygiidae) had different associations with small-scale habitat gradients inside *vs* outside reserves. These

patterns were consistent with a behavioural risk effect, where small fishes may be more strongly attracted to refuge habitats to avoid predators inside vs outside reserves.

The approaches developed and implemented in this thesis respond to some of the major current statistical and logistic challenges inherent in the analysis of counts of organisms. This work provides useful exemplar pathways for rigorous study design, modelling and inference in ecological systems.

Preface

I acknowledge the generous financial support of the Department of Conservation (Project Inv 4238), and Massey University's Institute for Natural and Mathematical Sciences (INMS) for providing a scholarship and, ultimately, a job. There are many people at Massey I wish to thank, including various office mates (Insha Ullah, Olly Hannaford, Rina Parry, Helen Smith, and Ting Dong), our wonderful administrative staff (Annette Warbrooke, Freda Mickisch, Lyn Shave, Anil Malhotra, Colleen Keelty (Van Es), and Vesna Davidovic-Alexander), and my INMS colleagues, particularly in the statistics and ecology groups, and my next-door-office neighbours, David Aguirre and Libby Liggins. Thanks also go to Assoc. Prof. Ann Dupuis for her advice and support through this process.

I could not have wished for a better primary supervisor than Professor Marti Anderson. Thank you, Marti, for always believing in me, despite often being confronted with compelling reasons not to. There were moments, in the latter stages, when you knew just the right thing to say to inspire me to grab this thing by the appendix and wrestle it into submission. You have given me so much and taught me so much about so many things but, foremost, thank you for being, in your words, my sternest critic and ultimately my biggest fan. Thanks also to my co-supervisor Russell Millar for keeping me on the methodological straight and narrow. Thank you to my late-coming co-supervisor, Matthew Pawley, for many reasons, but especially for the daily laughs and near-annual overseas adventures. Rather ironically, you've helped me keep my sanity through all this. I look forward to working with you in future, especially now that I am technically no longer your subordinate.

Many people spent long hours underwater counting fish for this PhD, including Oliver Hannaford, Marti Anderson, Steve Hathaway, Severine Dewas, Paul Caiger, Clinton Duffy, Charles Bedford, Kirstie Knowles, Nick Macrae, Sietse Bouma, Dave Culliford, Caroline Williams, and Alice Morrison. Particular thanks go to Clinton Duffy of the Department of

Conservation for skippering the RV Tuatini, and commenting on various manuscript drafts. Also, thank you to Steve Hathaway for bringing me fame by putting me on TV and in a book, and having the audacity to dub me a “guru”.

More broadly now, I wish to thank the people who inspired me to pursue a professional career in statistical and ecological research, and supporting me when it began in 2002, namely Jennifer Brown (University of Canterbury), Ian Westbrooke, and Ian West. I also thank Clinton Duffy (Department of Conservation) for showing me the water from the trees and inspiring my conversion to the study of things marine (you’re next, mate). A warm thank you goes to my father, Dr Murray Smith, for passing to me a small fraction of his extraordinary talent for statistics, and for encouraging me to take it on, along with other good advice when I needed it. Being able to work with you at NIWA has been a highlight of my career.

On a more personal note, I now turn to my little family. Being part of this family is the greatest privilege of my life. To my exquisite wife, Heidi, I offer you an ocean of gratitude for your unwavering support and patience. You are amazing and I could not have done this without you. Finally, to my beloved children, Finley and Anna. I am so proud and honoured to be your father. I cannot say that you made this endeavour any easier, but you and your mother make it and everything else worthwhile. My masters thesis was dedicated to Heidi, for it was during my masters that she agreed to marry me. You two graced our lives during this PhD, and I wholeheartedly dedicate it to you.



Me, bombastically gesticulating to Marti's bemusement.
Poor Knights Islands. (Photo credit: Steve Hathaway).

Table of contents

Chapter 1.	General introduction.....	1
1.1	Direct and indirect effects of marine reserves.....	1
1.2	Challenges in evaluating the effects of marine reserves	3
1.3	Aims	7
1.4	Overview of chapters	9
Chapter 2.	A review of Bayesian generalised linear mixed models for ecological studies	13
2.1	Introduction	13
2.2	Bayesian statistics—the basics.....	15
2.3	Example: an observational study of a marine reserve.....	20
2.4	Generalised linear models	22
2.5	Analysis of variance and mixed-effects models.....	29
2.6	Model fitting.....	39
2.7	Model evaluation and selection.....	43
2.8	Concluding remarks	52
Chapter 3.	Sources of zeros in ecological abundance data (Prologue to the study of snapper—Chapters 4 and 5)	54
3.1	Introduction	54
3.2	Zero counts in ecology	54
3.3	Excess zeros and the occupancy-abundance relationship	58
3.4	Zeros in counts of snapper from baited underwater video surveys.....	62
3.5	Concluding remarks	70
Chapter 4.	Incorporating the intraspecific occupancy-abundance relationship into zero- inflated models	72
4.1	Abstract	72
4.2	Introduction	73

4.3	The linked zero inflation model	76
4.4	Example.....	77
4.5	Discussion	83
4.6	Conclusion.....	87
4.7	Acknowledgements	87
Chapter 5.	Effects of marine reserves in the context of spatial and temporal variation: an analysis using Bayesian zero-inflated mixed models.....	89
5.1	Abstract	89
5.2	Introduction	90
5.3	Materials and methods	92
5.4	Results	101
5.5	Discussion	108
5.6	Acknowledgements	115
Chapter 6.	Marine reserves indirectly affect fine-scale habitat associations, but not density, of small benthic fishes	117
6.1	Abstract	117
6.2	Introduction	118
6.3	Methods.....	122
6.4	Results	128
6.5	Discussion	139
6.6	Acknowledgements	143
Chapter 7.	Could ecologists be more random?	144
7.1	Abstract	144
7.2	Main text	144
Chapter 8.	General discussion.....	154
8.1	Ecological effects of marine reserves.....	154
8.2	Statistical methodology	159
8.3	Summary	166

Appendix A Supplementary Material for Chapter 4.....	198
A.1 Formal description of linked zero-inflated negative binomial model	198
A.2 Table of summary statistics for estimated parameters	201
A.3 Potential relationships λ and π under linked zero-inflation.....	202
A.4 R and OpenBUGS code and data	203
A.5 Convergence diagnostics.....	205
A.6 Posterior predictive checks.....	208
A.7 Sensitivity Analysis.....	212
Appendix B Supplementary Material for Chapter 5	219
Appendix C Supplementary Material for Chapter 6	220
Appendix D Supplementary Material for Chapter 7	221
D.1 Table of useful spatial functions in R.....	221
D.2 Code for implementing random sampling designs.....	222
Appendix E Contribution to co-authored chapters	223

List of tables

Table 4-1. A comparison of a selection of candidate models for estimating the counts of legally sized snapper from a marine reserve monitoring program. For all models shown here, the base distribution for the counts was the negative binomial. Four classes of zero-inflated models were used, as indicated by the model numbers: (1) no zero inflation, (2) constant zero inflation, (3) a separate linear predictor for zero inflation, and (4) zero inflation linked to the mean of the count process. In the case of model 3, submodels 3.1–3.4 contain increasing numbers of parameters in the separate linear predictor for zero-inflation, as indicated. The predictor variables are denoted as follows: R = reserve status; S = season; A = area; Y = year. Models were compared using the Deviance Information Criterion (DIC) and its summands, the expected deviance (D) and the effective number of parameters (p_D). The actual number of stochastic parameters (p) is also provided. The mean of the posterior predictive distributions for the total number of zeros (Total n_0) and the total count (Total t) is presented. These may be compared with the same values from the observed data, namely 191 and 660, respectively. Finally, estimates of the mean absolute error for each of n_0 and t , pooled at the level of replicate bins, provide the “mean bin misclassification rate” (Bin ϵ_{n_0}) and the “mean bin absolute deviation” (Bin ϵ_t). For these measures, smaller values indicate more accurate predictions. (See Appendices A.1 and A.6 for further details of the model and posterior predictive checks.)81

Table 5-1. Details regarding the age and size of each of the three marine reserves examined in this study.93

Table 5-2. The number of baited underwater video (BUV) sampling units obtained in each year, season and location. Samples within each survey were allocated to reserve and non-reserve areas equally in most cases.93

Table 5-3. Sources of variation for the full ANOVA model, based all factors in the study design. The terms that were not included as candidates for model selection, based on preliminary heuristics, are indicated with an asterisk. The abbreviation for each term, as shown, was used to indicate the model parameters associated with that term in the GLMs, given in Equations (5-3) to (5-5) in the text. Terms that were chosen to be included in the final models of relative densities of legal or sublegal snapper, obtained using model selection on the basis of the DIC, are also provided.97

Table 5-4. Point estimates (mean of the posterior distribution, represented by the set of values given by MCMC) and 95% credible intervals (0.025 and 0.975 quantiles of the posterior distribution) of the mean relative densities for either sublegal or legal snapper in reserve and non-reserve areas at each of three locations. Reserve and non-reserve densities for sublegal snapper were pooled because there was no reserve effect in the model. Estimates of the ratio of reserve to non-reserve densities are also provided for legal snapper as an index of the ‘reserve effect’. The point estimates for the ratios were obtained by first calculating the ratios for each MCMC iteration, taking the natural log of the ratios, calculating the mean, and then back-transforming. 102

Table 5-5. Point estimates and 95% credible intervals (as described in the legend for Table 5-4) of the mean relative densities for either legal or sublegal snapper in each of two seasons at each of three locations. Estimates for ratios of seasonal effects were obtained as described for reserve effects in the caption for Table 5-4. 105

Table 6-1. The habitat features (and definitions) for which presence or absence was recorded in each quadrat. Also provided here are the loadings (standardised coefficients) of each habitat feature for each of the first three principal component axes from a PCA of 12 habitat variables. Loadings that exceed 0.4 are shown in bold. 124

Table 6-2. Taxa detected in the surveys, including the number of individuals of each taxon observed in the whole dataset. Most taxa were consistently identified to species, including all triplefins (TF). Moridae consisted mostly of the species *Pseudophycis breviuscula* and *Lotella rhacina*. Gobiesocidae were *Dellichthys morelandi* or *Gastrocyathus gracilis*. *Acanthoclinus* spp. were most likely the species *A. rua*, *A. marilynae*, or *A. littoreus*. The nine most abundant species are indicated by an asterisk (*) and these were modelled individually. 131

Table 6-3. Permutational multivariate analysis of variance (PERMANOVA) tests for the effects of variables on the structure of assemblages of benthic reef fish, based on Bray-Curtis distances of the transect-level abundance of all species. There were 635 residual *df*, and tests were based on Type III partial sums of squares and 999 permutations. 132

Table 7-1. Systematic review of methods used to choose spatial sampling units in recent ecological studies. We conducted a census of articles which resulted from a search on 7 July 2013 in Biological Abstracts. Search parameters were Year=2013, Topic=((abundan* OR densit*)), Major Concepts=(ECOLOGY), and Source Titles=(*Diversity and Distributions, Ecological Applications, Ecology, Ecology Letters, Journal of Applied Ecology, Journal of Animal Ecology, Journal of Ecology, Oecologia, or Oikos*). We only included studies that involved observations or experiments in the field, using spatially-replicated sampling units. Field experiments from a single block which was then divided into subplots were excluded. The allocation of treatments to units in experimental studies was not considered—we were only interested in the method used to choose the spatial locations of units. A total of 99 out of 215 articles met our criteria for review. 147

Table A-1. Prior distributions for stochastic parameters. 199

Table A-2. Summary statistics of the posterior distributions of estimated parameters, including the mean, standard deviation (SD), median, and 95% credible intervals (CI). 201

Table A-3. Summary of posterior predictive checks, by way of comparison of five summary test quantities calculated from the data ($T(\mathbf{y})$) with the posterior predictive distributions of the same test quantities, calculating from 5,000 replicate datasets simulated from the model ($T(\mathbf{yrep})$). Q 0.05 and Q 0.95 give the 90% credible intervals for the posterior predictive distributions. 210

Table A-4. List of models that were compared with the base model (1-*IZNB*) in a sensitivity analysis, with a coded description of their error structure as follows. Linked, Separate, and Constant zero inflation are indicated by LZI, SZI, and CZI, respectively (see Chapter 4 for definitions). For linked zero inflated model, the zero-inflation probability $\boldsymbol{\pi}$ was fitted as a function of the mean of the count distribution $\boldsymbol{\lambda}$, specifically, $f(\boldsymbol{\pi}) = \boldsymbol{\gamma}_0 + \boldsymbol{\gamma}_1 \log(\boldsymbol{\lambda})$, with f being the logit or the cloglog function, as indicated. Distributions used to model count values were Poisson (P), negative binomial (NB), or Poisson lognormal (PLN). The way in which each model differed from the base model is also explicitly described. Prior distributions used in the base model (for parameters whose priors were modified in other models herein) were as follows: $\boldsymbol{\beta}_0, \boldsymbol{\beta}_S, \boldsymbol{\beta}_R, \boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1 \sim N(0, 100)$; $\boldsymbol{\sigma}_A, \boldsymbol{\sigma}_Y \sim \text{half-Cauchy}(0,1)$; $\boldsymbol{\delta} \sim \text{Gamma}(10^{-4}, 10^{-4})$. An estimate and precision of the log reserve effect (shown as mean **LRE** and standard deviation \mathbf{s}_{LRE} of the posterior distribution) is given for each model.214

List of figures

- Figure 2-1.** Three alternative ways to graphically display an ANOVA design. The two dendrographs are a popular way of depicting an ANOVA design, but are not ideal for designs with both crossed and nested factors. The top dendrograph demonstrates that factors M and L are crossed by linking each level of one to each of the other, but it is then implied that there are only nine Areas in total, which is incorrect. Alternatively, the lower dendrograph shows that each level of L is replicated and achieves the correct number of Areas, but this visually implies that L is nested in M, as opposed to being crossed. The table logically displays the structure of the design while clearly indicating where factors are crossed or nested. Each cell represents a mean that is estimated.21
- Figure 2-2.** An illustration of overdispersed counts in a spatial context, by way of contagion and excess zeros. The points in panel A are randomly distributed without contagion or excess zeros; thus, counts taken from the underlying squares would not be overdispersed (the variance would equal the mean). Note that panel C contains a greater number of empty cells than panel A, illustrating that more zeros can arise through contagion alone without any explicit process that produces excess zeros (Warton 2005).25
- Figure 2-3.** Comparison of estimates of group means for treating the grouping factor as (a) fixed and (b) random, made by two models of the same dataset, plotted here on the same scale. The coloured circles at the bottom level represent the observed data, comprising three observations from 8 groups, which are connected by dotted lines to the estimates of their corresponding group means (coloured squares). For the fixed effect, there is no distribution fitted to the group means—they are simply estimated as the mean of the comprising values. For the random effect, a normal distribution was fitted with an estimated variance component, represented in grey. This results in more conservative estimates of the group means—they are shrunk toward the global mean. For both models, the errors from the group means were assumed to be normally distributed with an estimated error variance (as represented coloured distributions).34
- Figure 2-4.** Density of the Cauchy distribution with scale-parameter values of 1–3 (x-axis truncated at 8).35
- Figure 3-1.** A schematic diagram of a spatial intraspecific occupancy-abundance relationship (OAR). The locations of individuals in the study domain are shown in A. The shade of the small cells in B indicates the relative abundance of individuals per cell; white squares indicate non-occupancy. The OAR is illustrated in C at the spatial scale of the larger squares in B (each comprised of 6×6 cells); the number of occupied cells is positively related to the mean abundance per cell (figure reproduced from Borregaard and Rahbek 2010, with permission from Univ. of Chicago Press).59

Figure 3-2. The baited underwater video apparatus as used in the study of snapper (from Willis and Babcock 2000).....64

Figure 3-3. The occupancy-abundance relationship for legal (A) and sublegal (B) snapper, represented by the log conditional mean of the counts (λ_i) plotted against the logit probability of excess zero (π_i) estimated for each combination of area-by-year. The plotted numbers indicate the areas, and colours indicate inside (red) vs outside (black) marine reserves. These estimates came from zero-inflated models in which π_i and λ_i were fitted using separate linear predictors of the factors.68

Figure 4-1. The relationship between the conditional mean count (λ) of snapper (per baited underwater video deployment) and the probability of an excess zero (π) for legally sized snapper from a marine reserve monitoring program. This relationship was estimated using a Bayesian zero-inflated model (Appendix A.1) where π and λ were linked explicitly as $\text{logit}(\pi) = \gamma_0 + \gamma_1 \log(\lambda)$. The black line shows this function using point estimates for the parameters of $\gamma_0 = 0.34$ and $\gamma_1 = -1.60$. The grey lines show this function using the paired values of these parameters under MCMC within their joint 95% credible bounds.....82

Figure 5-1. A map showing the locations of three marine reserves in north-eastern New Zealand (upper left panel). Also shown are the individual numbered areas (fine lines and numbers), and marine reserves (bold lines) at each location, as indicated. Note that the borders of Tāwharanui Marine Reserve were moved slightly in September 2011 and are now different to those shown here.94

Figure 5-2. A variance components plot (Gelman 2005) showing the variation associated with each term in the chosen models, expressed as the estimate of the standard deviation σ among levels, for predicting the relative density of legal or sublegal snapper. For the latter, separate linear predictors were used to model the probability of an excess zero (π) and the conditional mean of the counts (λ), so a separate panel is used for each. Point estimates (means of posterior distributions) are represented by vertical lines, with 50% and 95% credible intervals for the means as thick and thin horizontal lines, respectively..... 103

Figure 5-3. Fine-scale spatial patterns in the estimated mean relative density of sublegal (triangles) and legal (circles) snapper, in areas within three locations. Open and closed symbols represent the point estimates (means of posterior distributions) for spring and autumn, respectively. Error bars are 95% credible intervals for the means. 106

Figure 5-4. Inter-annual and season patterns in the estimated mean relative density of sublegal (triangles) and legal (circles) snapper at three locations. Open and closed symbols represent the point estimates (means of posterior distributions) for spring and autumn, respectively. Error bars are 95% credible intervals for the means. For legal snapper, estimates for within the reserves only are shown, because too few snapper were observed outside the reserves to show any interpretable patterns. Note that the scale of the y-axes varies differ for sublegal (left) and legal (right) panels. 107

Figure 6-1. Predicted interactive effects of marine reserves (where a generalist predator is more abundant) and habitat complexity (where more complex habitat provides greater refuge from predation) on the densities of prey fish, under two different mechanisms. The two lines represent prey densities inside (in orange) and outside (in blue) marine reserves. In (a), the primary mechanism causing the interaction is predation, particularly in areas of low-complexity habitat. In this case, a main effect of reserve status is expected, where overall mean densities are lower inside reserves. In (b), the primary mechanism is a risk effect where prey fish, due to the abundance of predators, prefer more complex habitat in order to avoid predation. Here, no main effect of reserve status is expected and the average heights of the two lines are equivalent. 121

Figure 6-2. Map of the three locations and 35 sites. 123

Figure 6-3. Boxplots showing the distribution of values of depth and the three PCA axes for habitat for reserve and non-reserve transects at each location. 129

Figure 6-4. Non-metric multidimensional scaling (MDS) plots of (a) the Site-by-Year centroids and (b) Location-by-Reserve-by-Year centroids, in Bray-Curtis space, shown by Location and Reserve status (R = reserve, NR = non-reserve). The 2-d stress for the MDS analyses were 0.13 (a) and 0.07 (b). 133

Figure 6-5. Estimated mean densities per 5-m² transect for each combination of Location, Reserve status, and Year, with 95% credible intervals, for (a) species richness, (b) total number of fish, and (c–k) each of the nine most-common species. Estimates are for a standardised median habitat at a depth of 10 m. In some cases, the y-axes for (e–k) are shown on a square-root scale for clarity. 134

Figure 6-6. Estimated main effects of marine reserves, shown here as means and 95% credible intervals of the log ratios of Reserve vs Non-reserve means. They are shown for the overall study (i.e. calculated from reserve and non-reserve means that were averaged, or ‘marginalised’, across all Location-Year combinations), for each Location (averaged across Years), and for each Year (averaged across Locations). Filled circles indicate that the 95% credible interval does not contain zero, suggesting a non-zero difference associated with Reserve status. These estimates are standardised for habitat and depth using Bayesian hierarchical models. 135

Figure 6-7. Coefficients associated with the habitat axes, represented by the means and 95% CIs of the posterior distributions of coefficient values, estimated for the overall study and for each Location. Symbols for which the CI does not contain zero are filled, suggesting evidence for a non-zero habitat association. 136

Figure 6-8. Reserve-Habitat interactions, represented by the means and 95% CIs of the posterior distributions of the differences between the habitat coefficients inside vs outside reserves, estimated for the overall study and for each Location. Symbols for differences for

which the CI does not contain zero are filled, representing evidence for a non-zero interaction. 137

Figure 6-9. Exploration of Reserve-Habitat interactions, represented here as mean densities across the range of values of the habitat variables, estimated separately for inside and outside reserves, for the overall study and for each Location. Hab1 represents a gradient of increasing complexity, and Hab2 represents a gradient of sandy broken-up reef to solid reef with closed Ecklonia canopy. The y-axes are shown on a square-root scale for clarity. 138

Figure 7-1. Three simple sampling designs for selecting 20 sites (pluses) from within Tāwharanui Marine Reserve (orange) off Tāwharanui Peninsula (grey), near Auckland, New Zealand. All three designs were implemented in R by applying the function `spsample` to a polygon object `Strata`, in which the reserve is represented by four strata equally spaced along the seaward border. The designs shown are (1) Simple Random Sample—points selected randomly through the entire reserve [`spsample(Strata, n=20, type="random")`]; (2) Systematic Sample—points selected on a regular grid [`spsample(Strata, n=20, type="regular")`]; (3) Stratified Random Sample—five points selected from each of the four strata [`lapply(Strata@polygons, spsample, n=5, type="random")`]. Note that, in the latter case, five sites were taken from within each of the four strata for simplicity. In practice, it may be more efficient to allocate samples proportional to the stratum areas, which are straightforward to calculate in R (Table 2). These designs were all implemented using only freely available data. See Appendix D for further details, including full R code and data. 151

Figure 7-2. A more complex sampling design for Tāwharanui Marine Reserve (see Figure 7-1) targeting specific habitat and depth in four strata. The sampling frame and population of interest here was rocky reef habitat at depths of five meters. Panel A shows three spatial objects, representing the marine reserve (`Strata`, orange polygon), rocky reef (`Reef`, blue polygon) and a 5-m contour (`Contour`, black line). Three steps were taken to obtain the sample, as follows: 1. Convert the contour into regular points [`Pts <- spsample(Contour, n=1000, type="regular")`]. Note that, for clarity, only 100 points are shown in Panel B. 2. The points that overlie reef (black dots, panel B) are selected as candidate points while those that do not (white dots) are discarded [`PtsR <- Pts[!is.na(over(Pts, Reef)),]`]. 3. Randomly select five points from the candidates in each stratum (pluses in panel C) [`PtsRS <- tapply(PtsR, over(PtsR, Strata)$Name, sample, size=5)`]. See Appendix D for further details, including full R code and data. 152

Figure A-1. Potential relationships between the mean of the count distribution (λ) and the probability of an excess zero (π) under the general form of the linked model, $\text{logit}(\pi) = \gamma_0 + \gamma_1 \log(\lambda)$. The four lines in each panel have a different value of the intercept (γ_0) and a common slope (γ_1), as indicated. A large negative slope ($\gamma_1 = -10$) permits very fast transition from complete zero-inflation to no zero-inflation (A). This may be useful if low numbers are uncommon, so that zeros dominate below a particular threshold value of λ . Smaller negative

values of the slope can give a range of curves for decreasing zero-inflation with increasing abundance (B, C, D). For $\gamma_1 = 0$ (E), the relationship disappears and the model reduces to a constant value of π , as in model 2. While presumably unusual in nature, positive relationships between excess zeros and mean abundance can be generated (F).202

Figure A-2. Density histograms of posterior distributions of key model parameters, including the log of the reserve effect (“log.res.effect”). Mean and median values are shown as green and red vertical lines, respectively.....206

Figure A-3. Trace plots of the three MCMC chains, each shown in a different colour, of key model parameters, including the log of the reserve effect (“log.res.effect”).206

Figure A-4. Convergence diagnostic plots for model parameters (produced by the function plot.bugs() from the R2OpenBUGS package for R). The two plots on the left show the 80% credible intervals for each of the three chains in a different colour, and the Brooks-Gelman-Rubin convergence diagnostic \hat{r} (= “R-hat”). The plots on the right show overall medians and 80% credible intervals for the parameters and the deviance. Parameter “r” here is the dispersion parameter δ207

Figure A-5. Comparison of sorted data values between the observed data \mathbf{y} (x-axis) and 5,000 replicate datasets \mathbf{yrep} simulated under the fitted model (y-axis). The points represent mean values across the \mathbf{yrep} for each corresponding ranked value of \mathbf{y} . The grey polygon represents the 5% and 95% percentiles of values for \mathbf{yrep} at the corresponding rank. E.g., the highest value in the observed data was 25, whereas the highest values in the \mathbf{yrep} datasets was on average 22 and typically ranged from 15 to 33.209

Figure A-6. Sensitivity analysis in which the estimated log of the reserve effect from the base model (linked zero-inflated negative binomial, IZINB) is compared with a range of alternative models. Values shown represent means, and 2.5% and 97.5% quantiles, of the posterior distributions.216

Figure A-7. Sensitivity analysis in which the estimated means of counts (i.e., relative density of snapper) from non-reserve vs reserve areas from the base model (linked zero-inflated negative binomial, IZINB) are compared with a selection of alternative models. Values shown represent means, and 2.5% and 97.5% quantiles, of the posterior distributions.216