Contents

PREFACE ..... **V**

CONTENTS ..... **VIII**

1 INTRODUCTION TO BAYESIAN STATISTICS, MCMC TECHNIQUES, AND WINBUGS **1**

1.1 Probabilities and Bayes' Theorem 1
1.2 Likelihood functions 3
1.3 Conjugate prior distributions 6
1.4 MCMC 9
1.4.1 Markov Chain *9*
*1.4.2 Transition rules** *10*
1.5 Using WinBUGS 16
1.6 Summary 21

2 ZERO INFLATED GLMM APPLIED TO BARN OWL DATA **25**

2.1 Introduction 25
2.1.1 Vocal begging behaviour of nestling barn owls *25*
2.1.2 Previous analyses of the owl data *26*
2.1.3 Prerequisite knowledge for this chapter *28*
2.2 Importing and coding the data 28
2.3 Data exploration 29
2.4 Overdispersion in the Poisson GLMM 34
2.4.1 Assessing overdispersion using Pearson residuals *34*
2.4.2 Assessing overdispersion using an observation level random effect term *35*
2.4.3 Simulation study demonstrating observation level random effect *35*
2.4.4 A GLMM with observation level random effect *40*
2.5 Why zero inflated models? 41
2.6 Implementing a Poisson GLM in WinBUGS 43
2.6.1 Converting vectors to matrices *43*
2.6.2 Data for WinBUGS *45*
2.6.3 Modelling code for WinBUGS *45*
2.6.4 Initialising the chains *46*
2.6.5 Parameters, thinning rate and length of the chains *47*
2.6.6 Starting WinBUGS from within R *47*
2.6.7 Assessing convergence of the chains *48*
2.6.8 Summarising the posterior distributions *49*
2.6.9 Pearson residuals *51*
2.6.10 WinBUGS versus GLM results *53*
2.7 Implementing a Poisson GLMM in WinBUGS 54
2.8 Implementing a zero inflated Poisson GLM in WinBUGS using artificial data 56
2.9 Application of ZIP GLMM in WinBUGS using the owl data 62
2.10 Using DIC to find the optimal ZIP GLMM for the owl data 64
2.11 ZIP GAMM for 1-way nested data 66
2.12 What to present in a paper? 66

3 A ROADMAP FOR ANALYSIS OF OVERDISPERSED SANDEEL COUNTS IN SEAL SCAT **67**

3.1 Sandeel otoliths and seal scat 67

3.2 Data exploration 68
3.3 GLM with a Poisson distribution 71
3.4 GLM with a negative binomial distribution 73
3.5 GAM with a Poisson distribution 75
3.6 GAM with a negative binomial distribution 76
3.6.1 Model and R code 76
3.6.2 Model validation of the negative binomial GAM 77
3.6.3 Model interpretation 78
3.7 Zero inflated GAM with a Poisson distribution 80
3.7.1 True and false zeros 81
3.7.2 The ZIP model 81
3.7.3 Result for the ZIP GAM 83
3.7.4 Result for the ZINB GAM 84
3.8 Final remarks 87
3.9 Discussion 88
3.10 What to present in a paper? 89

4 GLMM AND ZERO INFLATED POISSON GLMM APPLIED TO 2-WAY NESTED MARMOT DATA 91

4.1 Introduction 91
4.2 Data exploration and visualisation 93
4.2.1 Import the data and get a first impression 93
4.2.2 Zero inflation in the response variable 94
4.2.3 Number of missing values 95
4.2.4 Outliers 96
4.2.5 Collinearity 96
4.2.6 Visualisation of relationships between number of young and the covariates 98
4.3 What makes this a difficult analysis? 101
4.4 Which statistical techniques and software to apply? 102
4.5 Short description of the statistical methodology 103
4.5.1 Viewing it (wrongly) as a Gaussian linear mixed effects model 103
4.5.2 Viewing the data (potentially wrongly) as a Poisson GLM or GLMM 106
4.5.3 Viewing the data (potentially wrong) as a zero inflated Poisson GLM 107
4.5.4 Viewing the data (possibly correct) as a 2-way nested ZIP GLMM 108
4.6 Dealing with 25 covariates in a 2-way nested ZIP GLMM 109
4.7 WinBUGS and R code for Poisson GLM, GLMM, and ZIP models 112
4.7.1 WinBUGS code for a Poisson GLM 112
4.7.2 WinBUGS code for 1-way and 2-way nested Poisson GLMMs 116
4.7.3 WinBUGS code for ZIP GLM 121
4.7.4 WinBUGS code for 2-way nested ZIP Poisson GLMMs 126
4.8 Validating the 2-way nested ZIP GLMM 128
4.9 Interpretation of the 2-way nested ZIP GLMM model 133
4.10 Discussion 134
4.11 What to present in a paper? 136
Appendix A: Correlation between observations in GLMMs* 137
A.1 Correlations in a Poisson GLMM for 1-way nested data 137*
A.2 Correlations in a Poisson GLMM for 2-way nested data 138*
A.3 Correlations in a binomial GLMM for 1-way nested data 140*
A.4 Correlations in a GLMM binomial model for 2-way nested data 142*
A.5 Correlations in a ZIP model for 1-way nested data 144*
A.6 Correlations in a ZIP model for 2-way nested data 145*
A.7 Non-technical summary of the appendix 145
A.8 Example of correlations for the 2-way nested ZIP GLMM 146*

5 TWO-STAGE GAMM APPLIED TO ZERO INFLATED COMMON MURRE DENSITY DATA 149

5.1 Introduction 149
5.2 Sampling 149
5.3 Covariates 153

5.4 DATA EXPLORATION 154
5.4.1 Potential outliers in the bird data 154
5.4.2 Zero inflation of the bird data 154
5.4.3 Outliers in the covariates 155
5.4.4 Collinearity of the covariates 156
5.4.5 Visualisation of relationships between birds and covariates 159
5.4.6 Cruise and transect effects 161
5.4.7 Summary of the data exploration 162
5.5 GAMM FOR ZERO INFLATED AND CORRELATED DENSITIES 163
5.5.1 Brainstorming 164
5.5.2 Four model selection approaches 165
5.6 RESULTS OF THE FULL MODEL APPROACH 165
5.6.1 Results of the presence/absence GAMM 165
5.6.2 More detailed output for the binomial GAMM 167
5.6.3 Post-hoc testing 168
5.5.4 Validation of the binomial GAMM 170
5.5.5 Analysis of the presence-only data using a gamma GAMM 172
5.7 FITTING A GAMMA GAMM WITH CAR CORRELATION IN WINBUGS** 173
*5.7.1 Fitting a Gamma GLM in WinBUGS** 174*
*5.7.2 Fitting a ZIP GAM on the Common Murre counts in WinBUGS** 176*
*5.7.3 Fitting a ZIP GAM with residual CAR in WinBUGS** 180*
5.8 DISCUSSION 181
5.9 WHAT TO PRESENT IN A PAPER? 182

6 ZERO INFLATED SPATIALLY CORRELATED COASTAL SKATE DATA 183

6.1 INTRODUCTION 183
6.2 IMPORTING THE DATA AND DATA CODING 183
6.3 DATA EXPLORATION 184
6.3.1 Outliers 184
6.3.2 Collinearity 185
6.3.3 Relationships 187
6.3.4 Geographical position of the sites 188
6.3.5 Zero inflation in the species data 189
6.3.6 What makes this a difficult analysis? 190
6.4 ZERO INFLATED GLM APPLIED TO *R. AGASSIZI* DATA 191
6.4.1 What is the starting point? 191
6.4.2 Poisson GLM applied to R. agassizi data 191
6.4.3 NB GLM applied to R. agassizi data 192
6.4.4 Zero inflated Poisson GLM applied to R. agassizi data 193
6.4.5 Zero inflated negative binomial GLM applied to R. agassizi data 196
6.4.6 Model validation of the ZIP GLM applied to R. agassizi data 201
6.5 ZERO INFLATED GLM APPLIED TO *A. CASTELNAUI* DATA 203
6.6 ADDING SPATIAL CORRELATION TO A GLM MODEL 206
6.6.1 Who are the neighbours? 206
6.6.2 Neighbouring sites 210
6.6.3 The CAR model 210*
6.6.4 Applying CAR to simulated spatially correlated data 212
6.7 ANALYSIS OF SPECIES 3: *SYMPTERYGIA BONAPARTII* 218
6.7.1 Poisson or negative binomial distribution? 218
6.7.2 Adding spatial correlation to the zero inflated Poisson GLM 218
6.7.3 Stetting up the required matrices for car.proper 219
6.7.4 Preparing MCMC code for a ZIP with residual CAR correlation 222
6.7.5 MCMC results for ZIP with residual CAR structure 225
6.8 DISCUSSION 229
6.9 WHAT TO PRESENT IN A PAPER? 230

7 ZERO INFLATED GLMS WITH SPATIAL CORRELATION– ANALYSIS OF PARROTFISH ABUNDANCE 231

7.1 INTRODUCTION 231
7.2 THE DATA 232
7.2.1 Surveying 232
7.2.2 Response variable 234
7.2.3 Covariates 234
7.3 ANALYSIS OF THE ZERO INFLATED DATA IGNORING THE CORRELATION 238
7.3.1 Poisson and NB GLMs 238
7.3.2 ZIP and ZINB GLMs 239
7.3.3 Model selection for ZINB 243
7.3.4 Independence 244
7.3.5 Independence – model misspecification 245
7.3.6 Adding year to the ZINB? 245
7.3.7 Dive effect on the residuals 246
7.3.8 Adjusting the sample variogram for land between transects 246
7.3.9 Which model to choose? 249
7.4 ZERO INFLATED MODELS WITH A RANDOM INTERCEPT FOR `DIVE` 250
7.5 ZERO INFLATED MODELS WITH SPATIAL CORRELATION 250
7.5.1 Adding a residual correlation structure to the ZINB 250
7.5.2 ZINB with a residual CAR correlation in R 251
7.5.3 MCMC results 253
7.6 PREDICTIONS 254
7.7 DISCUSSION 255
7.8 WHAT TO WRITE IN A PAPER 255

8 ANALYSIS OF ZERO INFLATED CLICK-BEETLE DATA 257

8.1 INTRODUCTION 257
8.2 THE SETUP OF THE EXPERIMENT 257
8.3 IMPORTING DATA AND CODING 258
8.4 DATA EXPLORATION 259
8.4.1 Spatial position of the sites 259
8.4.2 Response and explanatory variables 260
8.4.3 Viewing the data as time series 261
8.4.4 Spatial and temporal patterns 262
8.4.5 Big trouble 263
8.4.6 Effects and interactions for the species data 265
8.4.7 Zero inflation 267
8.4.8 Direction 267
8.4.9 Where to go from here; zero inflated models? 268
8.5. ZIP GLM 268
8.6 ZIP GLM WITHOUT CORRELATION FOR THE SPECIES DATA 269
8.7 RESULTS FOR THE FEMALE – MALE DATA 274
8.8 DISCUSSION 276
8.9 WHAT TO WRITE IN A PAPER? 276

9 ZERO INFLATED GAM FOR TEMPORAL CORRELATED SPERM WHALE STRANDINGS TIME SERIES 277

9.1 INTRODUCTION 277
9.2 WHAT MAKES THIS A DIFFICULT ANALYSIS? 277
9.3 IMPORTING AND DATA CODING 278
9.4 DATA EXPLORATION 278
9.5 POISSON GAM WITH A SINGLE SMOOTHER IN WINBUGS 280
9.5.1 Fitting a Poisson GAM using gam from mgcv 280
9.5.2 Fitting the binomial GAM from Pierce et al. (2007) 282
9.5.3 Knots 283
9.5.4 Low rank thin plate splines 284
9.5.5 Mathematics for low rank thin plate splines 286

9.5.6 Bypassing the mathematics ... *288*
9.5.7 WinBUGS code for GAM ... *289*
9.5.8 Results ... *290*
9.6 POISSON GAM WITH TWO SMOOTHERS IN WINBUGS ... 293
9.7 ZERO INFLATED POISSON GAM IN WINBUGS ... 296
9.7.1 Justification for zero inflated models ... *296*
9.7.2 Underlying equations for the ZIP GAM ... *297*
9.7.3 R code for a ZIP GAM ... *298*
9.7.4 Results for the ZIP GAM ... *299*
9.7.5 Model validation of the ZIP GAM ... *301*
9.8 ZERO INFLATED POISSON GAM WITH TEMPORAL CORRELATION IN WINBUGS ... 302
9.8.1 CAR residual correlation in the ZIP GAM ... *303*
9.8.2 Auto-regressive residual correlation in the ZIP GAM ... *306*
9.9 DISCUSSION ... 308
9.10 WHAT TO WRITE IN A PAPER? ... 308

REFERENCES ... 309

CONTRIBUTORS ... XIV