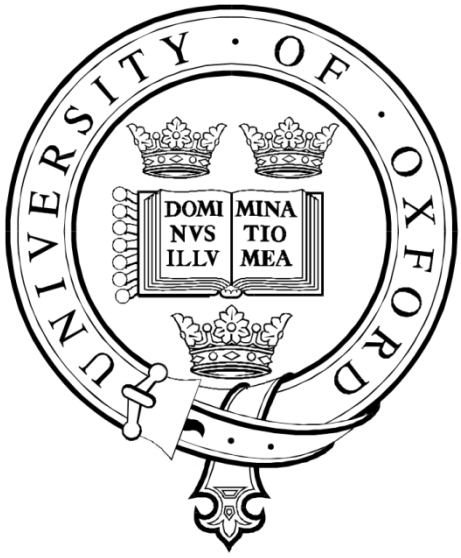
ANALYSING AND MODELLING THE IMPACT OF HABITAT FRAGMENTATION ON SPECIES DIVERSITY: A MACROECOLOGICAL PERSPECTIVE



Thomas J. Matthews

Linacre College
&
School of Geography and the Environment
University of Oxford

In fulfilment of the requirements for the degree of Doctor of Philosophy

Michaelmas Term 2014

ABSTRACT

Analysing and modelling the impact of habitat fragmentation on species diversity: a macroecological perspective

The destruction and fragmentation of natural habitat is the leading driver of the current extinction crisis. As a result, a research area has emerged focused on studying ecology in islands of natural habitat surrounded by a sea of anthropogenic land uses, so called 'habitat islands'. However, this research has largely been undertaken on small-medium scales, generally in single systems. Furthermore, many habitat island studies have used previous results and theory derived in the context of oceanic island research. Thus, this thesis aims to examine a variety of macroecological and biogeographical patterns using a large number of purely habitat datasets sourced from both the literature and my own sampling, with the objective of teasing out any general statistical patterns. The various macroecological and biogeographical patterns can be grouped under four broad headings: 1) species–area relationships (SAR), 2) nestedness, 3) species abundance distributions (SADs), and 4) species incidence functions (function of area). Overall, it was found that there were few hard macroecological generalities that hold in all cases across habitat island systems. This is because most habitat island systems are highly disturbed environments, with a variety of confounding variables and undesirable species (e.g. species associated with human land uses) acting to modulate the patterns of interest. Nonetheless, some clear patterns did emerge. In particular, it was found that the power model was by the far the best general SAR model for habitat islands. The slope of the island species–area relationship (ISAR) was related to the matrix type surrounding archipelagos, such that habitat island ISARs were shallower than true island ISARs. Significant compositional and functional nestedness was rare in habitat island datasets, although island area was seemingly responsible for what nestedness was observed. Species abundance distribution models were found to provide useful information for conservation in fragmented landscape, but the presence of undesirable species substantially

affected the shape of the SAD. Species incidence function curves were found to vary across space for a given species, but not in any systematic manner. In conclusion, this thesis finds that the application of standard island biogeography theory to habitat islands is too simplistic. It is hoped that the results of this thesis will act as a catalyst for a more macroecological approach to habitat island research, and a starting point for the development of an integrated theory of habitat islands.

EXECUTIVE SUMMARY

Ecologists have been studying habitat islands for over two decades and there are many published datasets that have yet to be brought together in any macroecological synthesis. In an attempt to fill this research gap, this thesis aims to identify any general macroecological and biogeographical rules in habitat islands using multiple datasets. The identification of any general rules will (i) aid in the development of a macroecological and biogeographical theory of habitat islands, and (ii) provide useful information for conservation in patchy landscapes. During a comprehensive search of the literature I sourced over 200 habitat island datasets, representing a wide variety of island types (e.g. forest fragments, grassland patches, and mountaintop islands) and taxa. These datasets were combined with four datasets I collected myself through two seasons of fieldwork, sampling birds in forest fragments in four different fragment landscapes across Europe. Further data were sourced from colleagues, including a well specified dataset of arthropods in forest fragments of the Azores, collected over a twenty year period.

Using this collection of habitat island datasets I was able to examine a wide variety of different macroecological and biogeographical patterns. These can be grouped under four broad headings:

1. **Species–area relationship (SAR):** the SAR has been a cornerstone of biogeographical research for over a century. However, our knowledge of the functional form of SARs in habitat island systems is still rather patchy. I examined a number of SAR related questions. These included: what is the best model/functional form of island SARs (ISARs) in habitat islands? Is there a threshold habitat island area below which the form of the ISAR changes? Do habitat generalist species act to modulate the shape of the ISAR?

2. **Nestedness patterns:** I was able to undertake a comprehensive evaluation of nestedness in the roughly 100 sourced datasets that included full species lists. Nestedness is an important component of conservation biogeography, because it influences policy, such as strategic conservation planning: highly nested systems require different conservation strategies to highly anti-nested systems. However, the prevalence of significant nestedness in habitat islands is largely unknown. Recent methodological advances have enabled workers to test for different types of nestedness, including functional nestedness (i.e. nestedness analysis of species similarity in terms of their functional traits). In this thesis, I examined both compositional and functional nestedness patterns, with a view to elucidating the role of island area in driving any observed nestedness.
3. **Species abundance distributions (SADs) hypotheses:** the extensive nature of the Azorean arthropod dataset allowed me to test various hypotheses regarding species abundance distributions (SADs) in fragmented landscapes. Namely, whether undesirable species (e.g. introduced species, and those species associated with human land uses) act to change the shape of the SAD. Furthermore, using the results from my various SAD analyses, I was able to argue the case for an increased uptake of SAD models in applied ecological research.
4. **Species' incidence function analysis:** using my fieldwork datasets, I attempted to answer the question of whether a species' incidence function (i.e. a function specifying the probability of occurrence of species at any island area value) is a constant trait across the species' range, or if there is some sort of systematic variation in incidence functions across the range. Based on the core-abundance hypothesis, I hypothesised that species should be more tolerant of fragmentation

at the range core relative to the edge, such that in core landscapes, species would have a higher incidence probability in smaller fragments.

Overall, my results indicate that it is difficult to define neat well specified generalities that hold in all cases. This is a characteristic of most ecological systems, but seems particularly problematic in habitat islands. Whilst in most instances it was not possible to explicitly test this hypothesis, it appears that a number of confounding variables affect the shape and form of macroecological patterns in habitat islands. In many ways this is to be expected as by definition most habitat island systems are highly disturbed anthropogenic environments. Future studies are required to establish conclusively the role of these various confounding variables, and to determine the extent to which they modulate the patterns examined in this thesis. Nevertheless, some clear patterns did emerge. Convex models, specifically the power model, are by far the best performing ISAR models, considering my datasets. The *z* value of the power model is related to island type, and was found to increase from habitat islands to oceanic islands, and within habitat islands with increasing hostility of the surrounding matrix. There was not much evidence of thresholds in the ISAR. Undesirable species (e.g. invasives and generalists) were found to affect both the slope of the ISAR, and the shape of the SAD. Significant compositional nestedness was relatively rare, although island area was important in driving the compositional and functional nestedness observed. Finally, the incidence functional analyses were inconclusive. For certain species, the hypothesised range effect was apparent; that is, for any given small fragment, there was a higher probability of presence in the range core landscapes than in the range edge landscapes. However, the opposite pattern was apparent for certain species, and thus the hypothesised range effect cannot be considered universal.

CONTENTS

Abstract **i**
Executive summary **iii**
Contents **vi**
List of tables **xiii**
List of figures **xv**
Acknowledgements **xviii**
Preface **xx**

CHAPTER 1: Introduction **1**
1.1 Conceptual framework 1
1.2 Research aims 4

CHAPTER 2: Literature review **7**
2.1 Habitat islands, habitat loss and habitat fragmentation 7
2.1.1. Habitat islands 7
2.1.2 Habitat loss and fragmentation 10
2.2 Macroecology and biogeography 35
2.2.1 Definitions 35
2.3 The species–area relationship 36
2.3.1 A brief history 36
2.3.2 A mélange of curves 38
2.3.3 Mechanisms driving the SAR 39
2.3.4 Biological interpretation 41
2.3.5 Applied uses of the SAR 43
2.4 Habitat amount thresholds 45
2.4.1 Types of threshold 45
2.4.2 Evidence for habitat amount thresholds 46
2.5 Nestedness 49
2.6 Metapopulations 52
2.7 Incidence functions 53
2.7.1 A brief history 53
2.7.2 The independent variable in incidence function models 56
2.7.3 Fitting standard incidence function (area) models 57
2.7.4 Applied uses of incidence function models 58
2.8 Macroecological patterns of species abundance 63
2.8.1 The species abundance distribution 64
2.8.2 Correlation between abundance and range size 80

CHAPTER 3: Methods **87**
3.1 Preamble 87
3.2 Data collection from the literature 87

3.3 Field site and patch selection 91
3.4 Patch characteristics 97
3.5 Bird sampling methodology 98
3.5.1 Sampling method 98
3.5.2 Target species 101
3.6 Additional site information 105
3.7 Bird trait data collection 108
3.7.1 Continuous traits 109
3.7.2 Forest specialism 109
3.8 Azorean arthropod data 109

CHAPTER 4: Thresholds and the species–area relationship: a synthetic analysis of habitat island datasets 112
4.1 Preface 112
4.2 Abstract 113
4.3 Introduction 114
4.4 Methods 118
4.4.1 Data collection 118
4.4.2 Data analysis 119
4.4.3 The effects of the dataset characteristics 122
4.5 Results 122
4.5.1 Model performance 122
4.5.2 Dataset descriptors 124
4.6 Discussion 127
4.6.1 Data transformation 127
4.6.2 Interpreting discontinuous models 131
4.6.3 Ecological interpretation 132
4.6.4 Conservation and methodological implications 132

CHAPTER 5: On the form of species–area relationships in habitat islands and true islands 136
5.1 Preface 136
5.2 Abstract 137
5.3 Introduction 138
5.4 Methods 140
5.4.1 Habitat island data collection 140
5.4.2 Model comparison 141
5.4.3 Model performance consistency 143
5.4.4 Explaining variation in ISAR patterns 144
5.5 Results 146
5.5.1 Model comparison 146
5.5.2 The power (log–log) model 149
5.6 Discussion 152
5.6.1 Model performance 153

5.6.2 Interpreting the parameters of the power (log–log) model 154
5.6.3 Conservation implications and conclusions 157

CHAPTER 6: Island species–area relationships and species accumulation curves are not equivalent: an analysis of habitat island datasets 161
6.1 Preface 161
6.2 Abstract 162
6.3 Introduction 163
6.4 Methods 166
6.4.1 Data collection 166
6.4.2 Data formatting and SAR construction 167
6.4.3 Relative influence of the predictor variables 169
6.4.4 Simulation analyses 171
6.4.5 Determining the degree of deviation of the archipelagic point 172
6.5 Results 173
6.5.1 Differences in the slope of ISARs and SACs 173
6.5.2 Boosted regression tree results 173
6.5.3 Nestedness simulations 177
6.5.4 The fit of the archipelagic data point 178
6.6 Discussion 179
6.6.1 Why are SACs generally steeper than ISARs? 179
6.6.2 Interpreting variation in ArcRes 182
6.6.3 Implications and conclusions 183

CHAPTER 7: Habitat fragmentation and the species–area relationship: a focus on total species richness obscures the impact of habitat loss on habitat specialists 185
7.1 Preface 185
7.2 Abstract 186
7.3 Introduction 187
7.4 Methods 189
7.4.1 Data acquisition and species classification 189
7.4.2 Data analysis 191
7.5 Results 193
7.5.1 Power (log–log) 193
7.5.2 Power (Non-linear) 195
7.5.3 Multi-model inference 196
7.5.4 Specialist:generalist ratio 198
7.6 Discussion 199
7.6.1 Generalist species mask the decline in specialists following habitat loss 199
7.6.2 Ratio of specialists to generalists 203
7.6.3 Conservation implications 204

CHAPTER 8: Quantifying and interpreting nestedness in habitat islands: a synthetic analysis of multiple datasets 206
8.1 Preface 206
8.2 Abstract 207
8.3 Introduction 208
8.4 Methods 214
8.4.1 Data collection 214
8.4.2 Quantifying nestedness 215
8.4.3 Determining the significance of nestedness patterns 216
8.4.4 Does island area underpin observed nestedness in habitat islands? 217
8.4.5 Habitat specialization 218
8.4.6 Model selection and dataset characteristics 218
8.4.7 Minimum set problem 219
8.5 Results 220
8.5.1 Prevalence of nestedness in habitat islands 220
8.5.2 Mechanisms and dataset characteristics 222
8.5.3 Minimum set results 224
8.6 Discussion 226
8.6.1 Nestedness prevalence and mechanisms 226
8.6.2 Are habitat specialists more nested than generalists? 228
8.6.3 Conservation implications 229

CHAPTER 9: Assessing functional nestedness patterns in habitat islands: an analysis of multiple bird datasets. 232
9.1 Preface 232
9.2 Abstract 233
9.3 Introduction 234
9.4 Methods 241
9.4.1 Dataset collection 241
9.4.2 Trait data 241
9.4.3 Measuring functional nestedness 242
9.4.4 Sensitivity analyses 245
9.5 Results 245
9.5.1 The nature of the datasets 245
9.5.2 Functional nestedness by island area 246
9.5.3 Sensitivity analyses 248
9.6 Discussion 249
9.6.1 The prevalence of functional nestedness in habitat islands 250
9.6.2 Driver of functional nestedness 252
9.6.3 Implications for conservation and ecosystem functioning 253
9.6.4 Concluding remarks 255

CHAPTER 10: The gambin model provides a superior fit to species abundance distributions with a single free parameter: evidence, implementation and interpretation 257
10.1 Preface 257
10.2 Abstract 258
10.3 Introduction 258
10.4 Methods 261
10.4.1 Gambin distribution 261
10.4.2 A description of the gambin R package 262
10.4.3 Azorean arthropod data 263
10.4.4 Model comparison 264
10.4.5 Sourced datasets 265
10.4.6 Sample size and grain size 265
10.4.7 Comparison of alpha values between land use types 268
10.4.8 Testing the sensitivity of alpha to the binning method 269
10.5 Results 270
10.5.1 Model comparison and inspection of the fit 270
10.5.2 The effect of sample size 270
10.5.3 Land use gradient 274
10.5.4 Sensitivity to the binning method 275
10.6 Discussion 276
10.6.1 Goodness of fit and multi-model comparison 277
10.6.2 Model performance and sample size 278
10.6.3 The alpha parameter as an ecological metric 279

CHAPTER 11: Multimodal species abundance distributions: a deconstruction approach reveals the processes behind the pattern 282
11.1 Preface 282
11.2 Abstract 283
11.3 Introduction 283
11.4 Methods 287
11.4.1 Study site and sampling methods 287
11.4.2 Species classification 288
11.4.3 Statistical analyses 289
11.5 Results 294
11.5.1 Model comparison 294
11.5.2. Post model selection 297
11.6 Discussion 301
11.6.1 A deconstruction approach 302
11.6.2 Distinguishing between two and three modes 307

CHAPTER 12: Examining variations in incidence functions across the range 310
12.1 Preface 310
12.2 Introduction 310

12.3 Methods 311
12.3.1 Data collection 311
12.3.2 Statistical analysis 312
12.4 Results 314
12.5 Discussion 317
12.6 Conclusions 322

CHAPTER 13: Discussion 324
13.1 Conservation actions in patchy landscapes 324
13.1.1 Strategic conservation planning 324
13.1.2 The assemblage deconstruction approach 328
13.1.3 The impact of the matrix on ISAR z values 330
13.1.4 Issues with incidence functions 331
13.2 Are there any macroecological generalities in regards to patchy landscapes? 331
13.2.1 Species abundance distributions 332
13.2.2 Species–area relationships 334
13.2.3 Nestedness 337
13.3 The differences and similarities between true islands and habitat islands 338
13.4 Key limitations 341
13.4.1 Limitations regarding the synthetic analyses 341
13.4.2 Limitations regarding my field work data 344
13.5 Future research opportunities 347
13.5.1 General research lacunae 347
13.5.2 Evaluating additional methodological approaches 349

CHAPTER 14: Conclusions 362

REFERENCES 366

APPENDIX 1: Neutral theory and the species abundance distribution: recent developments and prospects for unifying niche and neutral perspectives 398

APPENDIX 2: Supporting information: Thresholds and the species–area relationship: a synthetic analysis of habitat island datasets (Chapter 4) 436

APPENDIX 3: Supporting information: On the form of species–area relationships in habitat islands and true islands (Chapter 5) 452

APPENDIX 4: Supporting information: Island species–area relationships and species–accumulation curves are not equivalent data structures: a synthetic analysis of habitat island datasets (Chapter 6) 482

APPENDIX 5: Supporting information: Habitat fragmentation and the species–area relationship: a focus on total species richness obscures the impact of habitat loss on habitat specialists (Chapter 7) **494**

APPENDIX 6: Supporting information: Quantifying and interpreting nestedness in habitat islands: a synthetic analysis of multiple datasets (Chapter 8) **507**

APPENDIX 7: Supporting information: Assessing functional nestedness patterns in habitat islands: an analysis of multiple bird datasets (Chapter 9) **526**

APPENDIX 8: Supporting information: The gambin model provides a superior fit to species abundance distributions with a single free parameter: evidence, implementation and interpretation (Chapter 10) **528**

APPENDIX 9: Supporting information: Multimodal species abundance distributions: a deconstruction approach reveals the processes behind the pattern (Chapter 11) **542**

APPENDIX 10: Authorship statements for co-authored papers. **545**

APPENDIX 11: Dispensation letter for exceeding the allowed thesis word count. 556

LIST OF TABLES

Table 2.1: The effects of habitat destruction on biodiversity. 23
Table 2.2: The traits which dictate species sensitivity to fragmentation. 29
Table 2.3: The various proposed hypotheses predicting SARs. 40
Table 2.4: Descriptions of the various species abundance distribution models. 66
Table 2.5: The history of multimodal SAD distribution research. 70
Table 3.1: The target species. 104
Table 3.2: The environmental variables recorded at each sampling site. 106
Table 3.3: The variables used to determine the patch habitat complexity. 107
Table 4.1: The best fit island species–area relationship (ISAR) model for 76 habitat island datasets 124
Table 5.1: The 20 ISAR models compared in the model selection. 142
Table 5.2: Constrained analysis of principal coordinates (CAP) results for analyses of factors explaining the amount of variation in the model selection profile, for 202 habitat island datasets. 149
Table 5.3: Parameter estimates for a set of the most parsimonious linear models, modelling (a) the z value of the power (log–log model) and (b) the value of the c parameter. 153
Table 5.4: Reported average and/or range of z values of isolate systems from a selection of island species–area relationship syntheses and meta-analyses. 155
Table 6.1: The relative contributions (%) of predictor variables for boosted regression tree models developed using cross-validation. 175
Table 7.1: Model fit and ANCOVA results for the 23 bird–habitat island datasets. 197
Table 7.2: The z parameter of the non-linear power SAR model and associated lower (LCI) and upper (UCI) 95% confidence intervals for the 23 bird–habitat island datasets. 198
Table 8.1: The different factors proposed to contribute towards nestedness or conversely to anti-nestedness in islands and habitat islands. 211
Table 8.2: The nestedness metrics employed in the present study. 216
Table 8.3: The correlation of the row orders of 9 habitat island presence/absence matrices ordered according to decreasing area. 223
Table 8.4: Parameter estimates of a set of parsimonious generalised linear models with Gaussian variance, for 97 habitat island datasets. 224
Table 9.1: A glossary of terms used in this chapter. 236
Table 9.2: traitNODF and standard NODF values. 248
Table 9.3: TraitNODF area results from the sensitivity analysis. 249
Table 10.1: Additional datasets used in the multi-model comparison and the associated dataset characteristics. 267
Table 10.2: Goodness of fit and model selection results for arthropod SADs of 18 (No.) native Laurisilva forest fragments in the Azores. 272
Table 10.3: Wilcoxon rank sum test results for the significance of difference of α values of the gambin distribution between transects of different land use groups, for three Azorean islands. 276
Table 11.1: Site information for the different native forest fragments from the Azores used in the analyses. 293
Table 11.2: Site information for the different islands used in the analyses. 293
Table 11.3: Model selection results for arthropod SADs of twelve fragments of native Laurisilva forest in the Azores. 295
Table 11.4: Model selection results for arthropod SADs of five Azorean islands. 297

Table 11.5: Welch's Two Sample t-test results for the composition of the different modes within the SAD. 299
Table 12.1: The results of the global GLM analysis. 315
Table 12.2: The critical island area values. 315
Table 12.3: The combined binomial GLM results. 316
Table A1.1: Neutral theory and species-abundance distribution terminology. 404
Table A1.2: The different speciation modes that have been incorporated into neutral models. 410
Table A1.3: A summary of the developments in the history of the zero-sum multinomial distribution. 412
Table A2.1: Comparison of the number of selected best island species–area relationship models before (BC) and after (AC) applying the Cook's Distance criterion. 446
Table A2.2: The results of the sensitivity analysis: the best fit ISAR model for 76 habitat island datasets. 447
Table A2.3: The best fit ISAR model for habitat island datasets according to the three taxonomic groups. 448
Table A3.1: Description of the source papers. 453
Table A3.2: The results of the robustness analysis for 40 habitat island datasets. 473
Table A3.3: Parameter estimates for a set of the most parsimonious linear models modelling. 475
Table A4.1: The different types of species–area relationship. 483
Table A4.2: Description of the source papers. 485
Table A5.1: Description of the source papers. 494
Table A5.2: The *c* parameter and R^2 values of the power ISAR function for 23 bird habitat island datasets. 500
Table A5.3: Model selection results for all 23 bird–habitat island datasets. 501
Table A6.1: Datasets characteristics. 507
Table A6.2: The NODF values for the maximally packed matrix. 518
Table A6.3: NODF by rows and columns. 521
Table A6.4: The correlation of the row orders of 97 habitat island presence/absence matrices ordered according to decreasing island area. 523
Table A6.5: The NODF Z-scores (of the maximally packed matrix) calculated separately for habitat generalist and habitat specialist bird species. 524
Table A7.1: Description of the source papers. 526
Table A7.2 FF null model results. 527
Table A8.1: Site information for the 18 samples. 530
Table A8.2: Site information for the 6 Azorean island samples. 531
Table A8.3: Simulated lognormal metacommunity results. 532
Table A8.4: Simulated logseries metacommunity results. 534
Table A8.5: Model selection results for arthropod SADs of six Azorean islands. 535
Table A8.6: The proportion of an assemblage/metacommunity required to be sampled. 536
Table A8.7: Model comparison results. 537
Table A8.8: Simulation results. 538
Table A9.1: Likelihood Ratio Test results. 542
Table A9.2 Model selection results. 543

LIST OF FIGURES

Figure 2.1 The varying scales of habitat islands. 9
Figure 2.2 A cleared area of Boreal forest in Norway. 10
Figure 2.3 Temporal perspective of changes in species composition following a fragmentation event. 27
Figure 2.4 Timeline of the decline in species richness following a fragmentation event. 33
Figure 2.5: An example incidence function plot. 58
Figure 2.6 Exemplar SAD model fits. 68
Figure 2.7 Niche-oriented species abundance distribution models. 73
Figure 2.8 Using the species abundance distribution to determine the effect of disturbance on biotic community structure. 76
Figure 2.9 SAD-based methods for determining the effect of disturbance in applied ecology. 79
Figure 2.10 The relationship between global abundance and global range size. 81
Figure 3.1: A map of Europe with my study landscapes highlighted. 92
Figure 3.2: The location of my four study landscapes (red square) within the relevant countries. 93
Figure 3.3: An example study site in France: a forest fragment isolated by an agricultural matrix. 94
Figure 3.4: A fragmented cork oak landscape in Andalusia, South Spain. 95
Figure 3.5: My sampling accommodation in a mixed Boreal forest, Norway. 97
Figure 3.6: The view from a point count in a forest fragment in France. 101
Figure 3.7: The range of the Eurasian blue tit (*Parus caeruleus*). 104
Figure 3.8: The great spotted woodpecker (*Dendrocopos major*) – one of my target species. 105
Figure 3.9: A Laurisilva forest fragment on the island of Terceira, Azores. 111
Figure 4.1: Schematic illustration of the alternative models used in the analysis, and their different forms. 117
Figure 4.2: Box plots displaying the number of habitat islands (top row), and area of the smallest habitat island (bottom row), for each model type across the three data transformations. 126
Figure 4.3: Percentage of models identified to be the best depending on the minimum number of habitat islands (sample size) of a dataset selected, for the three versions of each dataset (untransformed, semi-log and log–log). 127
Figure 4.4: Variation in the best model selection across two versions of the same dataset. 129
Figure 4.5: Variation in the best model selection across three versions of the same dataset. 130
Figure 5.1: The performance of 20 species–area relationship models fitted to 202 habitat island datasets. 147
Figure 5.2: Change in model performance with the number of habitat islands in a dataset. 148
Figure 5.3: Variation in the *z* parameter of the power (log–log) model. 151
Figure 6.1: Power model fits constructed from the same habitat island dataset. 174
Figure 6.2: Partial dependence plots for a couple of the most influential variables in two boosted regression tree (BRT) models. 176
Figure 6.3: The relationship between nestedness and (a) zDif (b) the SAC *z* value, and (c) the ISAR *z* value. 177
Figure 6.4: The relationship between nestedness and the variation in *z* values of ISARs and small-large constructed SACs. 178

Figure 6.5: The island species–area relationship (power log–log model) for two habitat island datasets. 180
Figure 7.1: A selection of plots of ISARS using log–log power models, for six bird datasets. 194
Figure 7.2: Box plots displaying the z values of the power model (log–log)(a) and the non-linear power model (b). 195
Figure 7.3: A selection of derivative plots for three bird-habitat island datasets. 196
Figure 7.4: A selection of plots in which the ratio of specialist species to generalist species (S:G) against habitat island area is plotted for a selection of bird datasets. 199
Figure 8.1: A map of the 97 datasets (blue dots) included in this study. 221
Figure 8.2: Incidence plots for two habitat island datasets. 221
Figure 8.3: The relationship between the smallest proportion of sites required to represent all species in a dataset, and the NODF value. 225
Figure 9.1: A hypothetical example illustrating the concept of functional nestedness. 238
Figure 9.2: An exemplar dendrogram. 243
Figure 10.1: 1 Examples of the fit of the gambin distribution (black dots), the logseries (red triangles), and Poisson lognormal (blue squares) to observed data (bars). 273
Figure 10.2: Changing shape of the observed species abundance distribution with sample size. 274
Figure 10.3: Difference in the gambin distribution α parameter values (standardized to keep sample size constant) between transects of the three land use types (native forest, exotic plantation forest, and pasture) on three Azorean islands. 275
Figure 11.1: Map of study sites in the Azores. 288
Figure 11.2: Exemplar multimodal species abundance distributions. 300
Figure 11.3: The SAD of arthropods from the island of Faial in the Azores. 303
Figure 12.1: The incidence function plots of four target species. 317
Figure 13.1: Exemplar fits of the power (non-linear) species–area relationship model to six different data structures, constructed from the same habitat island dataset. 328
Figure 13.2: Variation in the z parameter of the power (log–log) model across habitat islands and true islands. 339
Figure 13.3: Model prediction uncertainty in ISAR extrapolation. 351
Figure 13.4: The results of a Bayesian metapopulation incidence function model fitted to a simulated dataset. 357
Figure 13.5: The posterior distributions of two logistic regression model parameters estimated using a Bayesian approach. 359
Figure A1.1: An illustration of Hubbell's (2001) classic two-tier spatially implicit neutral model. 407
Figure A1.2: Exemplar fits of three species abundance distribution models. 409
Figure A1.3: An illustration of how emergent neutrality can lead to multimodal species abundance distributions. 422
Figure A2.1: A selection of plots depicting discontinuous piecewise models selected as the best for particular datasets. 451
Figure A3.1: A map of the 207 datasets (blue dots) included in this study. 471
Figure A3.2: Variation in the *c* parameter of the power (log–log). 476
Figure A3.3: Variation in the *z* parameter of the power (log–log) model across habitat islands and true islands. 477
Figure A3.4: Variation in the *z* value of the power (log–log) species–area relationship across different area scales. 478
Figure A3.5: Variation in the *z* parameter of the power (log–log) model. 479
Figure A3.6: Variation in the *c* parameter of the power (log–log) model. 480

Figure A3.7: The functional form and first derivative of two ISAR models. 481
Figure A4.1: Exemplar fits of the power species–area relationship model. 484
Figure A4.2: The z value of the power model fitted using ISAR structured data plotted against z values derived using randomly constructed SAC data. 493
Figure A5.1: The derivative plots for all 23 bird–habitat island datasets. 503
Figure A5.2: The plots of the ratio of specialist species to generalist species. 505
Figure A6.1: The relationship between the number of species in a dataset (log transformed) and the NODF Z-score. 525
Figure A8.1: Observed and expected species according to the gambin distribution. 539
Figure A8.2: Difference in the logseries alpha parameter values (a-c) and the sigma parameter of the Poisson lognormal distribution. 540
Figure A8.3: The sensitivity of alpha to the binning method employed. 541

ACKNOWLEDGEMENTS

Firstly, I thank my parents for their considerable support during my DPhil study. I am particularly grateful to my Dad who has provided constant guidance and financial support throughout my academic career, as well as proof reading innumerable drafts. My Mum also kindly gave up many afternoons to help with data entry and proof reading.

My DPhil supervisor, Rob Whittaker, has provided invaluable support throughout the development of the thesis and has been central to the development of ideas, analyses and the various resultant publications. He has always found time to answer my questions and provide constructive feedback on ideas and drafts despite his multiple other commitments. He has also gone out of his way to provide me with opportunities to attend conferences, working group meetings, and to collaborate with researchers within his network, for which I am extremely grateful. I am also grateful to my undergraduate supervisor, Jon Sadler, for early encouragement and providing me with the opportunity to pursue a path in ecological research.

Special thanks are extended to those people who have collected and provided data that have been used in my thesis. In particular, I am grateful to Paulo Borges for allowing me to use the BALA dataset, which formed a central part of a number of my papers. I am also indebted to Paulo for offering me the opportunity to be an Associate PhD student in the Azorean Biodiversity Group, and for organising my trips to the Azores. Francois Guilhaumon and Carsten Rahbek also kindly arranged for me to visit and work at their respective institutions.

Numerous other researchers have provided guidance, feedback, methodological advice and the opportunity to collaborate on projects; without all of which my thesis would be a much reduced final product. There are too many to name here, but in particular, and in no particular order, I would like to acknowledge: Kostas Triantis, François Guilhaumon,

François Rigal, Karl Ugland, Michael Borregaard, Rich Grenyer, Shonil Bhagwat, Maan Barua, Eden Cottee-Jones, Alison Fairbrass, James Hale, Adam Bates, Paul Jepson, Lorraine Wild, Manuel Steinbauer, Zhe Sha, Luis Borda-de-Água, Catherine Sheard, Michael Athanson, Tom Bregman, and Andy Gosler. Eden Cottee-Jones kindly gave up several days to help practice point counts, identify obscure bird calls, and measure bird specimens. I am also very grateful to Alison Pool for her help in constructing many of the figures in the thesis, and associated papers and posters and for her patience in the face of my indecisiveness.

I am indebted to the numerous organisations that provided me with funding throughout my MSc and doctoral study at Oxford: the Royal Geographical Society (Post Graduate Award and Paddy Coker Award), the Sidney Perry Foundation, Santander, the EPA Cephalosporin Fund, the Sir Richard Stapley Trust, the Natural Environment Research Council, the John Edmonds' Trust, the Vice Chancellors' Trust, and Linacre College.

Finally, I thank Christie Palmer for her on-going support, motivation, and most of all, her enduring patience.

PREFACE

This thesis examines a wide variety of macroecological patterns using a selection of different habitat island datasets, and attempts to determine whether there are any macroecological and biogeographical rules and generalities which apply to habitat island systems. Habitat islands are defined more formally in Chapter 2, but I have taken the term to include a broad range of island types, including forest fragments, mountain top islands, and patches of grassland. The thesis combines meta-analyses of a large number of habitat island datasets with analysis of empirical data obtained through my own field work. The focus of the thesis is on multiple taxa, but my field work data pertains solely to birds. In addition to identifying any general rules, the thesis seeks to generate information and guidelines which can be used in the conservation and management of habitat island systems.

The thesis is structured according to the 'papers route', specified in the departmental guidelines. According to this structure, the results chapters are each individual standalone papers which have either been published or submitted to peer-reviewed journals, and are sandwiched between the standard thesis chapters, such as the methods and discussion. The papers cover a broad range of topics, and thus to aid interpretation a brief preface summarising the paper and the details of where it has been submitted is included before every results chapter. The first three thesis chapters constitute the introduction, literature review and methods section. The methods provided in Chapter 3 are an overview of the methods used, rather than providing specific information on particular data analyses; the latter are provided in the individual results chapters. Chapter 3 also outlines the methods employed during my own field work. The results are presented as individual papers and represent Chapters 4-12. Chapters 4-7, focus on species–area relationships and species accumulation curves, Chapters 8-9 cover nestedness and functional nestedness, Chapter 10-11 focus on species abundance

distributions, and finally Chapter 12 evaluates species incidence functions. The final two chapters (13 and 14) represent the standard thesis discussion and conclusion chapters.

All references are provided at the end of the thesis in a standalone section, i.e. the reference sections from the papers included as the results chapters have been removed. Whilst I acknowledge there are differences between the terms, I use macroecology and biogeography interchangeably throughout the thesis. In addition, the terms habitat islands, fragments, and patches are all used interchangeably.

The School of Geography and the Environment guidelines stipulate that thesis must be less than 100,000 words, including appendices. However, a number of published meta-analyses based on large numbers of datasets have been included in the thesis, and it would not have been possible to include them all, along with the relevant appendices, and still abide by the word count constraints. Thus, I have been awarded dispensation to submit a thesis in excess of this amount. The dispensation letter is provided as Appendix 11.

CHAPTER 1: INTRODUCTION

1.1 Conceptual framework

> [There is] an almost suicidal tendency for many ecologists to celebrate complexity and detail at the expense of bold, first-order phenomena. Of course the details matter. But we should concentrate on trying to see where the woods are, and why, before worrying about the individual trees.
>
> (Lawton, 1999, p.179)

The natural world is inherently patchy (Tilman & Kareiva, 1997), and most systems can be viewed as an island at some scale (Begon *et al.*, 1990; Whittaker & Fernández-Palacios, 2007). Island biogeography is the field of research interested in examining the factors that drive species diversity and species distributions in different types of islands. Whilst such a classification is evidently a simplification, the various types of islands studied in island biogeography research can be dichotomised, at least for terrestrial island types, into true islands, i.e. those surrounded by water (traditionally the main island type studied in island biogeography), and habitat islands (the focus of this thesis). The difference between true and habitat islands is somewhat subjective. Clearly, the archipelago of Hawaii can be classified as a set of oceanic islands, whilst South American Atlantic forest fragments surrounded by an agricultural matrix are clearly habitat islands. However, in certain other cases the distinction is not as clear. For example, increasingly many rainforest landscapes have been fragmented due to the creation of dams and reservoirs, resulting in the flooding of the surrounding landscape, and the creation of patches of forest surrounded by a matrix of water (e.g. Dalecky *et al.*, 2002; Yong *et al.*, 2011). Thus, strictly speaking these fragments are true islands in that they are surrounded by water (Whittaker & Fernández-Palacios, 2007), but in many ways the islands are more akin to habitat islands, e.g. in terms of the dominant assembly processes and sizes of the islands. For the purposes of this thesis, habitat islands are thus defined as any

island of natural habitat not surrounded by a marine water matrix; thus, excluding oceanic and continental shelf islands, but including flooded forest fragments. Regardless of the nuances in defining the different island types, it is clear that a field of research has emerged focused on studying biodiversity in habitat islands (Whittaker *et al.*, 2005). However, much of the work on habitat fragments and islands has been undertaken at the small to medium scale, focused on a single system, or a single or small group of species. As such, there have been calls for researchers to take a more macroecological perspective to the problem of habitat loss (e.g. Şekercioğlu & Sodhi, 2007; Prugh *et al.*, 2008). In this thesis macroecology is taken to refer to the analysis of the emergent statistical properties of ecological data sets.

The study of the macroecology of habitat islands is important for two reasons. First, from a theoretical perspective the development of a coherent theory of how species are distributed in habitat island systems is of great interest to macroecologists and biogeographers. The development of theory relating to oceanic islands has advanced ecology substantially over the past century, and benefited disciplines from evolutionary biology to dispersal ecology (MacArthur & Wilson, 1967; Williamson, 1981; Whittaker & Fernández-Palacios, 2007). A better understanding of species dynamics on habitat islands could have a similar effect (Harris, 1984; Ewers & Didham, 2006). Second, there is a practical necessity to studying habitat island systems from a macroecological perspective. The anthropogenic destruction, fragmentation and insularization of natural habitat are regarded as the primary causal factors driving the decline in global biodiversity (Wilcove, 1987; Sala *et al.*, 2000; Brook *et al.*, 2003; Schipper *et al.*, 2008). In many, if not most, terrestrial landscapes around the world, natural habitat now exists solely in patches of varying size, isolated by a matrix of various human land uses. The effects of this habitat insularization on species diversity and ecosystem functioning are still not fully understood (Fahrig, 2003; Lindenmayer & Fischer, 2006; Laurance, 2008). Thus, the identification of general macroecological patterns in habitat

islands can aid in the development of predictive tools of particular importance for conservation biogeography.

Studying diversity patterns in habitat islands from a macroecological perspective has a number of advantages. For instance, by using a large number of habitat island datasets it will be possible to elucidate any general patterns which are consistent across taxa and habitat island types; or to identify any systematic variation across taxa and island types. Workers have been studying individual habitat island systems for decades (e.g. Brown, 1971; Diamond *et al.*, 1987; Laurance *et al.*, 2002) and a large number of published datasets exist which have rarely been brought together in any macroecological synthesis. As Whittaker *et al.* (2005, p. 13) note, "numerous habitat island data sets have been collected, and so what is needed is not necessarily new field efforts, but more concerted efforts in analysis and synthesis, to tease out the scale sensitivity of habitat island data sets, particularly, the form of their [species–area relationship], and their compositional structure."

There is also a need to develop a theory of habitat island biogeography above and beyond that of oceanic island biogeography. Admittedly, I am not the first researcher to call for such a conceptual development (e.g. McIntyre & Hobbs, 1999; Haila, 2002; Lindenmayer & Fischer, 2006; Laurance, 2008), but arguably habitat island research is still studied through the lens of island biogeography theory derived from the study of mainly oceanic islands (e.g. MacArthur & Wilson, 1967). For example, most meta-analyses of macroecological patterns on islands have incorporated large numbers of oceanic island datasets (e.g. Patterson & Atmar, 1986; Ulrich & Gotelli, 2007; Triantis *et al.*, 2012); meaning that deriving any inferences of relevance to habitat islands is problematic. It has become increasingly realised that habitat islands differ from true islands in many ways (Watling & Donnelly, 2006; Laurance, 2008), e.g. in terms of the surrounding matrix and the main processes driving community assembly. Thus, one of the key questions in conservation biogeography is to what

extent do the various commonly studied biogeographical patterns (e.g. SARs, SADs, nestedness) differ between habitat and true islands, and how can any observed differences be linked to the aforementioned distinctions between the two types of island. As such, there is an urgent need for a thesis which aims to use appropriate statistical methods to investigate macroecological patterns, and the underlying processes, in habitat islands, and which synthesizes the findings in order to provide appropriate conservation and management guidelines (e.g. see Fischer & Lindenmayer, 2007).

1.2 Research aims

The overarching aim of my thesis is to develop a macroecological theory of habitat islands through an investigation of different biogeographical and macroecological patterns (and processes) in habitat island datasets. Through this endeavour, I aim to determine whether (a) there are any general rules regarding species distributions in habitat islands, and (b) it is possible to derive any information which can benefit biodiversity management and conservation.

To achieve these aims, four areas of research were pursued:

1) A comprehensive analysis of the relationship between species richness and area in habitat islands.

Whilst the island species–area relationship is one of ecology's oldest known patterns, the form of the relationship in habitat island systems is largely unknown. Chapter 5 compares the fit of 20 different ISAR models to 207 habitat island datasets, with an aim to determine (a) the form of the relationship in habitat islands, (b) the differences in the form of the ISAR between habitat and true islands, and (c) what dataset characteristics underpin variation in the parameters of the power model. Chapter 7 builds on this approach to determine how the form of the ISAR varies between forest specialist and forest generalist bird species.

Various different types of SARs have been documented, which can be dichotomised into island species–area relationships (ISARs) and species accumulation curves (SACs). Chapter 6 evaluates the differences between ISARs and SACs using 97 habitat island datasets, and assesses how well the standard power log–log ISAR model is able to predict the richness of larger areas.

2) The prevalence of taxonomic and functional nestedness in habitat islands.

Most nestedness meta-analyses have included large numbers of oceanic and continental shelf island datasets. Thus, the prevalence of significant nestedness in habitat island archipelagos is largely unknown. Chapter 8 uses currently accepted appropriate statistical and null modelling methods to determine what proportion of 97 habitat island datasets are significantly nested, and to what extent island area drives any observed nestedness.

Recent nestedness publications have introduced novel nestedness metrics which enable workers to search for nestedness patterns in types of diversity other than species composition (i.e. taxonomic nestedness). Using a set of bird-habitat island datasets and a comprehensive functional traits database, Chapter 9 aims to determine to what extent habitat island systems are functionally nested.

3) Using the species abundance distribution as a tool to derive useful information for applied ecology in habitat island systems.

There has been a recent resurgence of interest in SAD over the last decade. However, much of this work has focused on the theoretical aspects of the pattern, with less focus on using the SAD for applied purposes. Chapter 10 introduces a novel method for fitting the gambin SAD model (Ugland *et al.,* 2007), and illustrates how gambin can provide useful information regarding the shape of SADs in fragmented and disturbed landscapes. Chapter 11 examines the prevalence of multimodal SADs using a well specified dataset of arthropods in forest

fragments in the Azores, and in a similar vein to the analyses in Chapter 10, aims to use multimodal SAD models to generate information that is of use in applied contexts.

4) Assessing whether species incidence function models are an effective tool to investigate the effects of habitat insularization of species' distributions.

Species incidence function models have been argued to represent useful tools for guiding species-specific conservation decisions in fragmented landscapes. However, the extent to which species' incidence functions vary systematically across space, and thus the utility of concepts such as minimum area requirements and critical island area values, is largely unknown. Chapter 12 compares the incidence functions (incidence as a function of island area) of six species across four fragmented landscapes in Europe to determine to what degree species' incidence functions vary across a species range. Unlike the remainder of the chapters, this analysis is based solely on original field data collected specifically for the purposes of the thesis.

CHAPTER 2: Literature Review

2.1 Habitat islands, habitat loss and habitat fragmentation

2.1.1. Habitat islands

> In short, there can be few natural communities lacking at least some element of 'islandness'
>
> (Begon *et al.*, 1990, p. 768)

Islands have long been considered natural laboratories for ecological research due to the fact that they are discrete and quantifiable units and cover a wide range of abiotic conditions; thus proving researchers with ideal systems to test and develop ecological theory (Darwin, 1859; MacArthur & Wilson, 1967; Whittaker & Fernández-Palacios, 2007). However, until the late 1960s island biogeography, and island research more generally, was focused primarily on the study of true islands. In this thesis, true islands are defined as any set of islands surrounded by a matrix of marine water (i.e. continental fragments, continental shelf islands and oceanic islands; see Whittaker & Fernández-Palacios, 2007). During the 1970s, the extent of anthropogenic habitat destruction and fragmentation was beginning to be realised (see Harris, 1984). The process of habitat fragmentation results in the creation of isolated patches of natural habitat set within a matrix of human land uses (e.g. agriculture, urban land uses). These isolated patches came to be viewed as habitat islands; isolated in a sea of man-made habitats (Preston, 1962; MacArthur *et al.*, 1972; Simberloff & Abele, 1976). Thus, researchers began to apply island biogeography theory to the study of habitat islands, both from a theoretical and an applied perspective (Diamond, 1975a, 1976, 1984; Wilson & Willis, 1975; Simberloff & Abele, 1982; Harris, 1984; Brown & Dinsmore, 1988).

Early on, habitat islands became almost synonymous with true islands, with the same terminology being used to describe both island types, e.g. 'land bridge' was used to refer to a strip of land connecting a continental shelf island to the mainland, and to corridors of natural habitat connecting fragments to continuous habitat (MacArthur *et al.*, 1972). To take another

example, Harris (1984) discussed how habitat islands can be dichotomised into ‘fragments’, which are akin to continental shelf islands due to the fact that they both “start” with a full complement of species, and ‘mountaintop islands’ which are more akin to oceanic islands. However, whilst useful for conceptualising early habitat island research, this approach overlooked the obvious differences between true islands and habitat islands. Of these, the main differences are the permeability of the surrounding matrix (Ricketts, 2001; Prugh, 2008; Watling *et al.*, 2011), varying edge effects (Laurance & Yensen, 1991; Ewers & Didham, 2006), and in the case of oceanic islands, the dominant assembly processes (i.e. immigration vs. speciation; see section 13.3 in Chapter 13 for a more in depth discussion on the differences between true and habitat islands). The application of MacArthur & Wilson’s (1967) equilibrium island biogeography theory to habitat islands has also been criticised on the grounds that it ignores species interactions (e.g. predation), which are known to be important processes in habitat island systems (Laurance, 2008). For instance, the loss of top predators has been shown to result in mesopredator release in various urban habitat island systems (Soulé *et al.*, 1988; Crooks & Soulé, 1999). The concept of equilibrium in itself (e.g. MacArthur & Wilson, 1967) has also been criticised as being unhelpful in the study of habitat islands which are generally very dynamic systems (Pickett *et al.*, 1992; Whittaker & Fernández-Palacios, 2007). Finally, it has been argued that the island analogy is detrimental as use of the term island implies that a system is isolated from other influences, but most habitat island systems are set within human modified landscapes and are affected by a large number of external threats (Janzen, 1986).

The term ‘habitat island’ is quite vague and has been used to refer to a variety of different island types varying substantially in scale (see Fig. 2.1), including forest fragments (Marini, 2001), grassland fragments (Guardiola *et al.*, 2013), mountaintop sky islands (Brown, 1971), lakes (Browne, 1981), individual plants (Southwood & Kennedy, 1983; Davis

& Jones, 1986), and even pockets of air space (Diehl, 2013). What is immediately apparent when looking at these different habitat island types is that some are naturally fragmented systems (e.g. mountaintop islands) whilst others are fragmented as a result of anthropogenic habitat fragmentation (e.g. most forest fragments). There are many similarities between natural and human generated habitat islands, including the fact that the surrounding matrices are generally more permeable than the ocean (Harris, 1984). Nonetheless, there are differences, such as the time since isolation and the prevalence of external human driven threats, and it is important to keep this distinction in mind. I stated above that matrices are generally more permeable for habitat islands relative to true islands because this is not necessarily always the case, and arguably many habitat island systems may be effectively more isolated than certain true island systems. For example, mountaintop forest islands separated by large distances of desert are arguably more isolated than continental shelf islands separated by relatively small distances of salt water (Whittaker & Fernández-Palacios, 2007). Notwithstanding these differences, in this thesis, habitat islands are defined as any set of islands not surrounded by a matrix of marine water.



Figure 2.1 The varying scales of habitat islands: isolated trees, habitat fragments and whole mountain ranges. The left and centre photos were taken in France during my field work, and the photo on the right is of the Snake Range in the Great Basin National Park, USA, and is courtesy of the United States Geological Survey.

2.1.2 Habitat loss and fragmentation

> Habitat fragmentation is the most serious threat to biological diversity and is the primary cause of the present extinction crisis.
>
> (Wilcox & Murphy, 1985, p. 884)

As stated in the previous section, habitat islands are both a natural phenomenon and the result of anthropogenic habitat destruction and fragmentation (e.g. Fig. 2.2). This section is focused on the latter type of habitat island and provides an overview of the vast habitat loss and fragmentation literature.



Figure 2.2 A cleared area of Boreal forest in Norway.

2.1.2.1 Definitions: Habitat loss or habitat fragmentation?

While evidently inter-linked, habitat loss and fragmentation are distinct concepts. Habitat loss refers purely to a reduction in patch area without reference to the 'configuration' of the remaining habitat, while the process of fragmentation involves the subdivision of once continuous habitat, i.e. an increasing isolation of patches (Fahrig, 1998; Drinnan, 2005), often resulting in patches interspersed within a habitat matrix of human dominated habitats, such as

agriculture and urban land uses (Bender *et al.,* 1998; Ewers & Didham, 2006). In reality, the process of fragmentation generally incorporates a reduction in habitat patch area (Smith *et al.,* 2009; Mortelliti *et al.,* 2010). Confusingly, the term 'fragmentation' is frequently used in different contexts. It is used both to refer to i) the process of dividing up continuous habitat into smaller segments resulting in a reduction in area and an increase in isolation, and also to ii) the aforementioned pure 'habitat loss,' i.e. simply a reduction in habitat area, with no change in isolation (Franklin *et al.,* 2002). Due to this confusion, some workers have argued that the term 'fragmentation' represents a panchreston; too broad a concept to be of much use in a research context. To overcome this confusion, Hu *et al.* (2012) divided fragmentation into two types: fragmentation with no reduction in habitat amount, and fragmentation with a reduction in the amount of habitat.

As described above, fragmentation and habitat loss are distinct processes. However, it is not always a simple task to separate the effects of habitat loss (a simple reduction in patch area) from those of fragmentation (including increased patch isolation), and it has been argued that this often leads to erroneous predictions of the effects of fragmentation and increased isolation, that is, it is the reduction in area which drives, *inter alia,* species declines, rather than increased isolation (Fahrig, 1997; Johnstone *et al.,* 2014). The difficulty arises as both processes generally happen simultaneously in real landscapes, and are thus strongly correlated (Smith *et al.,* 2009; Johnstone *et al.,* 2014). Nonetheless, it is important to attempt to disentangle the relative influences of each process as they may require separate conservation measures (Betts *et al.,* 2006; Lindenmayer & Fischer, 2007). For instance, if a species is only sensitive to habitat amount and not habitat isolation, the protection of as much habitat as possible should be the primary conservation objective.

Yaacobi *et al.* (2007) attempted to isolate the effect of habitat fragmentation from pure habitat loss using species-area curves and Fisher's α diversity index, for plants and

carabids in a fragmented Mediterranean scrub landscape. The authors found that reductions in isolation did not influence species diversity when decoupled from the effects of a reduction in area; although, the authors go on to argue that isolation will significantly affect certain species, particularly those which require large expanses of primary habitat (Yaacobi *et al.*, 2007). This finding was echoed by Gavish *et al.* (2012) who compared four methods aimed at separating fragmentation effects from habitat loss effects, for spiders in an Israeli agricultural landscape. Interestingly, Gavish and others found that diversity actually increased with increasing fragmentation for all four of the methods employed. It is likely that the habitat matrix plays a large role in determining the effect of fragmentation with a permeable, less contrasting, habitat matrix, as in the two aforementioned studies, allowing many species to utilise resources within, and easily disperse across, the matrix and thus actually supplementing the available patch habitat (Brotons *et al.*, 2003). However, it is, therefore, important to remember the converse situation: fragmentation effects may be amplified if the surrounding matrix is highly contrasting to the patch habitat, e.g. urban land uses (Dunford & Freemark, 2005). As Yaacobi *et al.* (2007) stress, it is also important to recognise that certain taxa do not act in accordance with these generalisations, and particular species are highly sensitive to even small degrees of fragmentation, regardless of matrix type (see also Primack, 2010).

Villard & Metzger (2014) have recently argued that the importance of the configuration of habitat in the landscape is determined by the amount of habitat; a point of view which again highlights the linkages between habitat loss and fragmentation. When the amount of landscape is high, most species are specialists and there is low beta-diversity in the system. In contrast, when the amount of habitat is low, all species are generalists and there is low beta diversity. At intermediate habitat level amounts there are a mixture of species, beta-

diversity is high and the configuration of habitat is then critical (Villard & Metzger, 2014). This idea links into to concept of habitat amount thresholds, which are discussed below.

As a final note on definitions, whilst habitat loss and fragmentation are generally defined solely as spatial phenomena, they can also be considered temporally. For instance, prolonged periods of disturbance and unsuitable conditions for persistence of certain species within individual spatial units, such as the mowing of a grassland fragment, have been termed fragmentation in time (Poschlod *et al.,* 1998).

2.1.2.2 Extent

A large number of studies have attempted to estimate the rate and amount of habitat destruction. These studies have generally focused on tropical deforestation, and have produced substantially different estimates (e.g. Myers, 1986; Laurance & Bierregaard, 1997; Foley *et al.,* 2005). To take two credible examples, the average annual loss of tropical forest has been estimated at 5.8 million hectares (Laurance & Bierregaard, 1997; Laurance & Peres, 2006), and 40% of the Earth's land surface is now estimated to comprise cropland and pastures (Asner *et al.,* 2004; Foley *et al.,* 2005). Rather than focusing on the global figures, perhaps a more compelling approach is to provide examples of individual regions. For example, the Paraná State, Brazil, was comprised of almost continuous Atlantic Rainforest until the 1920s. However, clear felling of the forest for agricultural expansion has since reduced the forest cover to less than 1% of its former range, leaving behind small fragments which are increasingly isolated (dos Anjos, 2001).

Habitat loss and fragmentation are not random processes in terms of where they occur. Generally, within a given region it is the productive lowland areas which are converted to human land uses first (Liu & Silk, 2014). This can have important knock on effects as these are often areas with high biodiversity value (Margules & Pressey, 2000). In addition, the pattern of fragmentation has been found to be dependent on the agent of habitat

destruction. For example, Rosa *et al.* (2012) found that due to changes in Brazilian government policy, in the Amazon large continuous forest patches were generally ignored by loggers, and deforestation was focused primarily on smaller patches.

2.1.2.3 Perception

At this point, it is necessary to highlight the fact that most fragmentation research is conducted in landscapes which are viewed as fragmented from a human perspective. Equally, ecological studies of continuous habitats are based on human perceptions of what constitutes continuous; the same landscape may be highly fragmented from the perception of certain taxa (Wiens & Milne, 1989). The binary dichotomy of habitat and non-habitat matrix, central to many fragmentation frameworks (discussed below), suffers from this human perception issue as it overlooks the ecological concept of species-specific habitat definitions (Fischer & Lindenmayer, 2006). This can have important implications as it is easy to make erroneous conclusions based on imprecise identifications of what constitutes a habitat or what constitutes fragmented, for a given taxon. In this respect, it is important to consider how the taxon under study may perceive a fragment and/or fragmented landscape (Wiens, 1976; Wiens & Milne, 1989; Dunford & Freemark, 2005). For example, a bird species may be considered able to easily cross a small gap between forest fragments. However, if that bird species has evolved in dense forest over millions of years, there may a psychological barrier preventing it from crossing even very small distances (Dunford & Freemark, 2005).

2.1.2.4 Fragmentation frameworks

Habitat fragmentation, and habitat fragments, have been studied through a number of different conceptual frameworks (for a review see Lindenmayer & Fischer, 2006; Didham *et al.*, 2012). As stated above, the original approach to studying fragmented landscapes was through the lens of island biogeography theory (MacArthur & Wilson, 1963, 1967; for a review see Haila, 2002). During the 1980s and 1990s this approach was superseded by

landscape ecology, such as McIntyre & Hobbs' (1999) model of different fragmented landscape configurations, and metapopulation ecology (Hanski, 1999; see section 2.7 for a brief overview). Neutral theory (Hubbell, 2001; see Appendix 1) has also been shown to be a useful framework for studying the effects of fragmentation as it allows workers to isolate the importance of species abundance and dispersal, independently of species traits (Hubbell, 2001; Gilbert *et al.,* 2006).

All of the above models/frameworks are centred on the idea of patches set within a matrix of inhospitable habitat (Fischer & Lindenmayer, 2006). Such a view has been criticised on the grounds that it is grounded too heavily in humans' perceptions of the landscape (Manning *et al.,* 2004). More recently, a number of alternative frameworks have been proposed which have a greater focus on continuous heterogeneous landscapes, rather than the binary divide between habitat and matrix (Haila, 2002; Manning *et al.,* 2004; Fischer & Lindenmayer, 2006). These continuum models are based on Gleason's (1926) early concept of individualist species associations and the idea that species distributions change gradually along environmental gradients; that is, the landscape is viewed as a continuum of different habitats (Fischer & Lindenmayer, 2006). Inherent within these models are the assumptions that species differ in how they perceive the landscape, and that ecological processes affect habitat suitability in spatially continuous and elaborate ways (Fischer & Lindenmayer, 2006; Lindenmayer & Fischer, 2006). The continua-umwelt model of Manning *et al.* (2004) is a good example of such a framework. As Lindenmayer & Fischer (2006) note, there is in fact a spectrum of fragmentation frameworks, ranging from the binary perspective of equilibrium island theory at one end, to the full scale continua models at the other.

Recent conceptual advances include the habitat amount hypothesis of Fahrig (2013; see also Villard & Metzger, 2014) which postulates that landscape species richness is driven primarily by the amount of habitat in the landscape, irrespective of other landscape

characteristics, such as the size and configuration of the individual patches. Finally, a recent paper by Ewers *et al.* (2013) introduced the landscape 'terrageny' concept for fragmentation research. This approach involves representing the historical spatial separation of fragments in a similar way to how phylogenetic trees model evolutionary branching between species (Ewers *et al.,* 2013). The terrageny approach is interesting because it incorporates past landscape history, the so called 'ghost of land use past', which has been shown to be important in driving current species distributions. For instance, a number of studies have investigated potential extinction debts through determining whether past or current landscape patterns (i.e. the size and distribution of fragments) explain current species distributions (e.g. Krauss *et al.,* 2010; Guardiola *et al.,* 2013). The terrageny approach spawns a number of interesting predictions, such as a decline in community similarity with terragenetic distance.

2.1.2.5 A summary of fragmentation effects on biodiversity

The destruction and fragmentation of natural habitat are widely agreed to be the primary causes of the decline in global biodiversity (Wilcove, 1987; Sala *et al.,* 2000; Barbosa & Marquet, 2002; Primack, 2010). However, beyond this general statement, the exact impact of fragmentation on biodiversity is as yet unknown, and as Watson (2002, p.823) states, "our understanding of habitat fragmentation, is ironically, fragmented." For example, in a review of fragmentation experiments, Debinski & Holt (2000) concluded that there are no effects of fragmentation which are consistent across all systems and taxa. Despite this uncertainty, studies have reported myriad deleterious effects of fragmentation on species diversity patterns and ecosystem functioning, and I will now attempt to synthesise the results of these studies here. Table 2.1 provides a general summary of the various effects discussed. Mostly, the effects are reviewed individually, but it is important to recognise that generally the effects of fragmentation act in tandem with other drivers of species loss (e.g. climate change induced range shifts; Laurance & Bierregaard, 1997; Thomas, 2000; Robinson & Sherry, 2012). In

addition, it is necessary to stress that the negative effects of habitat loss and fragmentation extend further than direct impacts on species diversity; the loss of habitats is known to have wide ranging ecosystem effects, such as contributing towards climate change, and increased flooding and soil erosion (Foley *et al.,* 2005).

According to basic island theory, habitat loss and fragmentation (for the rest of this section, simply fragmentation) affect species through a reduction in habitat area and an increase in isolation (Whittaker & Fernández-Palacios, 2007). The initial habitat destruction extirpates any species that were endemic to the destroyed habitat (He & Hubbell, 2011). Any surviving species that are reliant on resources (e.g. habitat types, microclimatic conditions) that have now been fully lost are also consigned to extinction (Wilcove *et al.,* 1986). Extinction rates are increased as the sizes of populations in the remaining fragments are generally smaller than pre-fragmentation conditions. Small populations have an increased risk of extinction due to a number of factors, including Allee effects, and demographic, environmental and genetic stochasticity (Table 2.1; Gilpin & Soulé, 1986; Wilcove, 1987; Franklin & Frankham, 1998; Primack, 2010). The smaller a population gets the more vulnerable it is to further environmental, demographic and genetic stochasticity, and the population can be seen to decline further towards extinction, a process termed the extinction vortex (Gilpin & Soulé, 1986). The reduced genetic diversity present in small populations also limits the capability of populations to evolve and adapt to future changes and disturbances, further increasing the extinction risk of small populations (Harris, 1984). Due to these factors, there is generally considered to be a threshold population size, below which populations quickly decline towards extinction: the minimum viable population size (MVP; Reed *et al.,* 2003). The exact definition of the MVP is debated, but is generally taken as something like the population size required for there to be a 95% chance of survival for 1000 years (Whittaker & Fernández-Palacios, 2007). A review of MVPs is beyond the scope of the

current thesis, but they have been found to vary considerably amongst taxa (see Traill *et al.*, 2007). It has been the argued that the MVP should be viewed as a graded concept. For example, the 50:500:5000 rule states that an effective population size of 50 individuals is required to maintain short term fitness; whilst, 500 individuals is needed to balance the loss of genetic variation with that gained through mutation; and finally a population of over 5000 is required to maintain long term evolutionary potential (see Franklin & Frankham, 1998). Notwithstanding these arguments, the basic idea of MVPs in the context of fragmentation research is useful: that below a certain size, populations in fragments become too small to persist, and without rescue from immigrant individuals, the populations will go extinct.

According to standard island biogeography theory, the increased isolation resulting from fragmentation should reduce species richness in fragments due to a) reduced colonisation of patches, and b) a reduction in the rescue effect of small populations (Brown & Kodric-Brown, 1977; Hanski, 1999; Foppen *et al.*, 2000; Laurance *et al.*, 2002). Increased isolation also affects species in other ways, such as individuals' ability to find mates, and the reduced probability that all the resources a species needs to fulfil its life history have made it to a patch, e.g. pollinator species (Dick, 2001). However, whereas the effects of a reduction in habitat area have been convincingly documented for most taxa, the effects of isolation in fragmented landscapes tend to be much more species specific, with species responding to isolation in positive, negative or neutral ways (Bender *et al.*, 2003). Determining the relationship between isolation and species richness is complicated by the fact that isolation can be measured in many ways. The standard isolation metric employed is the Euclidean distance from a fragment to either the nearest fragment, or the nearest fragment above a certain size (Prugh, 2009). However, this approach ignores the characteristics of the interweaving matrix, an integral component of isolation (Bender & Fahrig, 2005; Watling *et al.*, 2011). More permeable matrices reduce the effective isolation of fragments for most

species; although not all, as for certain taxa even small, relatively permeable barriers between patches may represent insurmountable obstacles (Laurance *et al.*, 2002; Dunford & Freemark, 2005). As a result, studies have shown that isolation is more effectively measured using buffer type metrics, that is, metrics which incorporate the amount of habitat and non-habitat within a set buffer zone around a fragment (Bender *et al.*, 2003; Prugh, 2009).

Changes in species richness are only part of the story, and species abundance and/or density are also affected by fragmentation. Generally speaking, population sizes are theorised to decrease following a reduction in habitat area (MacArthur & Wilson, 1967), and this is indeed the case for many species (Gonzalez *et al.*, 1998; Didham *et al.*, 2012). In contrast, some species have been observed to become "hyper abundant", due to changes in system dynamics (Laurance *et al.*, 2002; Bennett *et al.*, 2014). The most well cited example of hyper abundance is the concept of mesopredator release, whereby the loss of predators from a region, due for example to resource demands which can no longer be met, results in an increase in the populations of lower level predators and/or herbivores (Crooks & Soulé, 1999; Terborgh *et al.*, 2001). This in turn can result in the reduced abundance of producers at the bottom of the food chain, leading to secondary extinctions (Wilcox & Murphy, 1985), and eventually to 'trophic and ecosystem collapse' (Terborgh *et al.*, 2001; Dobson *et al.*, 2006).

In regards to density, a number of theories have been applied to fragmented systems to make predictions about how density will change. For example, whilst population sizes are predicted to decrease with decreasing island size, the ETIB predicts that overall density will remain constant (MacArthur & Wilson, 1967). This prediction is known to be false, at least over short time scales, as following a fragmentation event species density increases, at least temporarily, as individuals pack into the remaining habitat (Laurance & Bierregaard, 1996). A contrasting theory to the predictions of ETIB is MacArthur *et al.*'s (1972) density compensation hypothesis. Applied to habitat islands, MacArthur *et al.*'s hypothesis postulates

that species present on both large and small habitat islands should have higher densities on smaller islands due to reduced interspecific competition resulting from lower species richness: zero sum dynamics of sorts. Empirical tests of density compensation in terrestrial systems have produced mixed results, with many studies finding little evidence that it occurs (e.g. Brotons *et al.*, 2003; Houlahan *et al.*, 2007), possibly due to compensatory effects of the surrounding matrix (Brotons *et al.*, 2003). The habitat compensation hypothesis postulates that the density of certain species will be greater in fragmented habitat, relative to un-fragmented habitat, because matrices often provides resources and additional habitat which can sustain larger populations of species than would otherwise be the case (Brotons *et al.*, 2003).

In addition to detrimental effects on traditional taxonomic aspects of diversity (e.g. species richness and abundance), fragmentation also impacts functional and phylogenetic diversity (Table 2.1; Ding *et al.*, 2013; Newbold *et al.*, 2013; Frishkoff *et al.*, 2014). For instance, Ding *et al.* (2013) examined the effects of fragmentation on bird diversity on landbridge islands in China using a suite of functional diversity metrics. It was shown that functional evenness decreased with decreasing island area, and thus highlighted the importance of examining aspects of diversity other than standard taxonomic diversity measures (Ding *et al.*, 2013). As sensitivity to fragmentation varies across taxa (discussed below), certain areas of trait space (e.g. large bodied species) are more likely than others to be lost in the fragmentation process (i.e. the loss of functional diversity is not random), which can reduce functional redundancy (Newbold *et al.*, 2013).

The aforementioned effects of habitat loss are those predicted by standard island biogeography theory, i.e. a decrease in habitat amount and an increase in isolation. However, fragmented landscapes differ from oceanic islands in that the surrounding landscape is not comprised of salt water; rather, the matrix surrounding habitat islands can comprise multiple

land use types. This fact means that the effects of habitat loss go beyond what is predicted by island theory. Although a gross simplification, these additional effects can be dichotomised into edge effects, and matrix permeability/connectivity effects. Each will now be discussed in turn.

The boundary between a habitat island and the surrounding landscape can influence species within the island in numerous ways. This can be in terms of altered flows, e.g. an increase in predators and matrix species entering the patch, or an increase in patch individuals dispersing out of the patch into habitat in which mortality rates are increased (Ries *et al.,* 2004). This altered flow regime can actually increase overall patch richness, as generalist and matrix species enter the patch, although this is not a universal finding (Ewers & Didham, 2006). Additional impacts include: i) an increased risk of fire (Cochrane & Laurance, 2002), ii) increased input of seeds from the surrounding landscape which changes the plant composition within the patch (Wilcove, 1987), iii) changes in species interactions near the boundary (e.g. Aizen *et al.,* 2012) and iv) changes in microclimatic conditions in habitat adjacent to the edge (e.g. increased wind speeds, increased light intensity, and increased temperature; Laurance *et al.,* 2002). That being said, edge effects are not easily predictable and studies have reported positive, negative and neutral edge effects depending on various factors, including the taxon under study, the type of edge, the severity of fragmentation and time of year (Ries *et al.,* 2004). Edge effects are also dynamic, and generally lessen with time as the gaps begin to 'seal' (Lovejoy *et al.,* 1986; Ewers & Didham, 2006). The overall impact of edge effects in a given patch also depends on the patch shape (Laurance *et al.,* 1998; Laurance, 2008). A circular patch has a higher core:edge ratio, relative to an oblong patch of the same area and thus the severity of any edge effects are likely to be lower in the circular patch (Ewers & Didham, 2007).

In many ways, the matrix composition determines the severity of any edge effects, as it is the matrix which defines the contrast at the edge. However, there are a number of factors related to the matrix over and above what can be described as edge effects. As mentioned above, the matrix acts as a source of multiple negative factors, including disease, invasive species, and fire (Janzen, 1986). Due to these factors, some habitat island matrices have been argued to be more hostile than the sea surrounding oceanic islands (Wilcove *et al.,* 1986). That being said, the matrix can have positive impacts for certain patch species. A relatively "friendly" matrix is likely to contain resources, and possibly even additional habitat, which patch species can utilise (Estades, 2001; Brotons *et al.,* 2003; Cook *et al.,* 2004). For example, certain species have been observed to increase in abundance following habitat loss, which has been put down to the additional resources which become available when access to the surrounding matrix increases (Estades, 2001). A non-hostile matrix can also result in increased inter-patch dispersal (due to both reduced dispersal mortality and an increased willingness of individuals to disperse from a patch) leading to a variety of rescue effects, e.g. an increase in genetic diversity (Ricketts, 2001; Laurance *et al.,* 2002; Haynes & Cronin, 2003). These benefits depend on both the individual species and the actual matrix composition: some species rarely leave the patch even when the matrix habitat seems favourable, whilst a very hostile matrix (e.g. heavy urban land uses) is unlikely to offer any of the aforementioned benefits. Thus, a number of authors have argued for an increased recognition of matrix conservation (e.g. Pereira & Daily, 2006; Ladle & Whittaker, 2011; Driscoll *et al.,* 2013). The argument being that in many human landscapes increasing the amount of habitat is not an option, but improvements to the matrix, e.g. environmentally friendly farming and the construction of corridors and hedge rows, are more feasible.

Table 2.1 The effects of habitat destruction and fragmentation on biodiversity.

Impact	Description	Sources
Decrease of the core: edge ratio	Leads to an increase in edge effects (discussed below) and a higher likeliness of disturbances with the patch.	Wiens (1995); Primack (2010)
Edge effects	An increase in temperature and wind speed and a decrease in soil humidity at the edges of habitat patches have been found to negatively affect forest specialist species richness. These types of conditions favour the matrix generalist species allowing them to increase in richness in the edge of patches. Laurance *et al.* (1998) found edge effects affect forests up to 60m in from the edge in the Amazon Rainforest. There is also an increase in inter-species interactions, namely between specialist patch species and those from the surrounding matrix (including humans), which increases predation rates and the spread of disease.	Laurance *et al.* (1998); Tabanez & Viana (2000); Primack (2010)
Matrix effects	The matrix can have positive and negative effects depending on the species. For some species, a non-hostile matrix provides additional resources to utilise, and can even provide additional habitat for species to exploit. However, matrices can also act as sinks, drawing species out from fragments into areas with increased mortality rates. Many external threats to fragment biotas originate in the matrix, including humans. Depending on the permeability of the matrix, inter-patch dispersal can be relatively easy for most taxa, or can result in high dispersal mortality.	Janzen (1986); Gascon *et al.* (1999); Laurance *et al.* (2002); Brotons *et al.* (2003); Driscoll *et al.* (2013)
Behavioural dynamics	The reduction in area leads to certain species being unable to carry out all stages of their life history that they could pre-fragmentation.	Tabanez & Viana (2000)
Increased competition	Some studies have argued that fragmentation will reduce competition due to generation of additional niche axes, and a reduced opportunity for any one species to dominate. In contrast, others have argued that as the pool of resources declines competition will be elevated.	Feeley (2003); Bregman *et al.* (*in review*)
Loss of ecosystem function and	A reduction in functional diversity, including the loss of whole functional groups, can have substantial effects on ecosystem functioning. This problem is accentuated as	Kruess & Tscharntke (1994); Wiens

phylogenetic diversity	fragmentation generally results in the loss of species with particular traits. Furthermore, important functions that humans rely on, such as biocontrol of agricultural pest species, have also been found to be negatively affected by fragmentation. Finally, evolutionarily distinct species have been found to have higher extinction rates in fragmented systems.	(1995); Crooks & Soulé (1999); Newbold *et al.* (2013); Frishkoff *et al.* (2014)
Loss of trophic levels	Species at the top of the food chain have larger metabolic demands, meaning they are often absent from fragmented landscapes due to insufficient resources. This in turn can result in an increase in the abundance of prey species, i.e. mesopredator release.	Wilson & Willis (1975); Crooks & Soulé (1999)
Changes in assemblage structure	As specific subsets of species are extirpated following a fragmentation event the patch assemblage becomes homogenous. This homogeneity arises as widespread and ubiquitous habitat generalists dominate at the expense of rarer habitat specialists.	Watson (2002)
Immediate extinction	Species may become extinct immediately following habitat loss because a) they were endemic to the destroyed habitat, b) they relied on resources which were only found in the destroyed habitat, or c) they have minimum area requirements which are no longer met.	Wilcove (1987); Ewers & Didham (2006); He & Hubbell (2011)
Reduction of population size (small population issues)	A reduction in population size can lead to genetic drift as rare alleles are lost. A further reduction in genetic variability can result from inbreeding depression, outbreeding depression and a loss of evolutionary flexibility. Small populations are also at greater risk of demographic stochasticity, i.e. natural variations in the birth and death rate, and environmental perturbations. The smaller a population gets the more vulnerable it is to further environmental, demographic and genetic stochasticity, and the population can be seen to decline further towards extinction, a process termed the extinction vortex.	Wiens (1995); Primack (2010)
Limits to dispersal	The increased isolation of fragments can affect metapopulation dynamics as well as restricting certain species' access to mates and resources. It also lowers the probability of the rescue effect saving species from extirpation.	Primack (2010)
Morphological changes	A variety of morphological changes in species exposed to fragmentation have been observed. Such changes occur through natural selection acting on species exposed to new conditions, and through stress. For example, it has	Schmidt & Jensen (2003); Bonte *et al.* (2012); Van

	been found that the body size of generalist rodents increased in a set of fragments due to the ability of these species to utilise resources in the patch and matrix habitats, and the increased resource availability resulting from the loss of specialist species. Factors such as road kill from increased traffic have also resulted in changes to average population body size in mammals in Denmark. Fragmentation also influences the costs of dispersal, and species will adapt and evolve in response to these changes in different ways; with some species committing evolutionary suicide.	Der Mesch *et al.* (2013)
Fluctuating asymmetry	A particular type of morphological change, characterised by differences between the left and the right characteristics of an individual that are normally symmetrical. These differences are caused by stress during development. Studies have discovered that fluctuating asymmetry is higher in birds present in smaller patches as opposed to larger patches.	Anciães & Marini (2000); Schmeller *et al.* (2011)
Disease	Following fragmentation the density of species often increases as individuals 'pack in' to the remaining habitat. This can increase the risk of disease transmission. The introduction of certain species (e.g. rats) into fragmented landscapes by humans increases the number of disease vectors. Finally, studies have observed an increase in human exposure to certain diseases, e.g. Lyme's disease, in fragmented landscapes.	Dobson & May (1986); Allan *et al.* (2003); Foley *et al.* (2005)
Loss of species' interactions	Fragmentation can result in the loss of many important species interactions, such as mutualistic networks. For example, a number of studies have shown that due to a loss of specialist species, fragmentation substantially alters pollination network architecture, which in time will result in a number of plant extinctions. However, invasive species have been found to take up the pollination roles of extirpated pollinators in fragmented landscapes.	Dick (2001); Fortuna *et al.* (2013); Spiesman & Inouye (2014)
Synergistic effects with other extinction drivers	There are many examples, but the interaction between fragmentation and climate change is the most studied. Fragmentation reduces species' ability to respond to climate change through range shifts by reducing available habitat and increasing the distance species need to migrate.	Hill *et al.* (1999, 2001); Ewers & Didham (2006)

2.1.2.6 Variation in the effects of fragmentation across taxa

2.1.2.6.1 Specialists and generalists

Fragmentation leads to significant changes in the proportions of habitat specialists and habitat generalists. Specialists are at particular risk from fragmentation as they generally have low abundances (Gaston, 1994), are heavily reliant on within patch resources which decline following a fragmentation event (Tellería & Santos, 1995), and possess "insuperable psychological barriers" preventing them from dispersing across gaps between patches (Diamond & May, 1976). As the numbers of specialists decrease following habitat fragmentation, they are replaced by ubiquitous generalist species that thrive in the resultant disturbed environment (Newbold *et al.,* 2013, 2014). As a general rule, generalist species are better adapted to surviving in, and utilising resources within, the matrix, and are thus well suited to fragmented landscapes (Brotons *et al.*, 2003; Bommarco *et al.*, 2010).

This shift in species composition is of particular conservation relevance as the influx of generalists from the matrix following fragmentation can 'mask' the decline in the once abundant specialists; this having the effect of depressing the slope parameter of various SAR functions (see Fig 2.3 for an illustration of this process, and how the relative proportions of specialists and generalists change through time) and obscuring the relationship between patch area and species of conservation concern, i.e. habitat specialists (Magura *et al.*, 2001; Pavlacky & Anderson, 2007). Despite the conservation implications that arise from this 'masking effect', the concept has attracted limited research. One interesting example, though, is a study by Magura *et al.* (2001). Magura *et al.* studied Carabids in a fragmented landscape in Hungary, and found that the SAR for the total species compliment was in fact significantly negative. However, when the SAR calculations were run separately for specialists the relationship became significantly positive, highlighting the statistical influence of including generalist species in calculations.

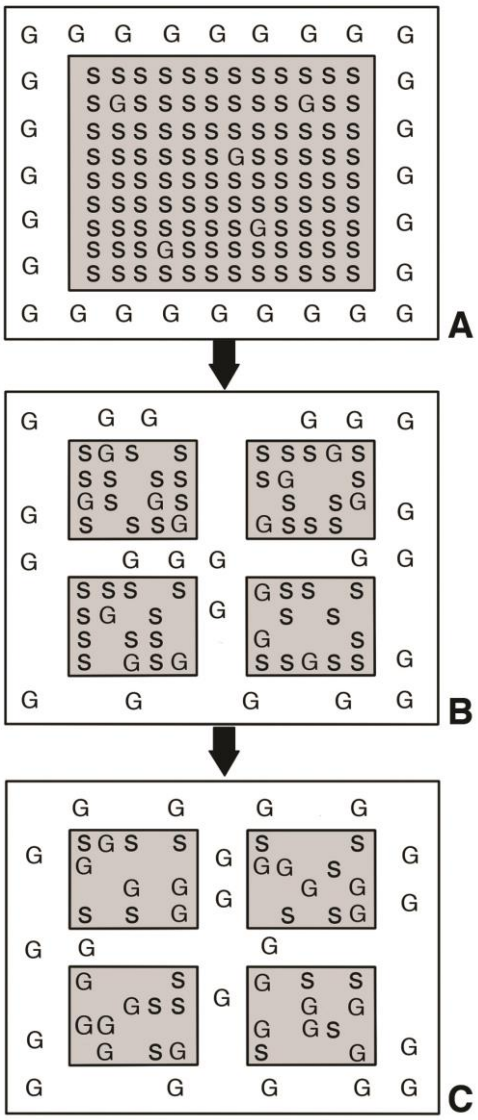


Figure 2.3 Temporal perspective of changes in species composition following a fragmentation event. S= habitat specialists; G = habitat generalists; Grey shading = forested areas; White = surrounding matrix. A) Illustrates a continuous forest containing a biota comprising of predominantly specialists (S) species. Immediately following a fragmentation event (B) a number of specialist species are lost as a direct result of habitat destruction, but the proportion of specialists in the fragment remains high. However, as time passes (C), there is an influx of generalist species (G) and the proportion of the specialists in the fragments declines. Based on concepts in Watson (2002).

2.1.2.6.2 Which species are sensitive to habitat loss?

Aside from specialists and generalists, there are substantial variations in the response to fragmentation across taxa (Terborgh, 1974; Henle *et al.,* 2004; Robinson & Sherry, 2012; Newbold *et al.,* 2013; Bregman *et al.,* 2014). This variation has been linked to a large number of different traits, and I have attempted to condense this literature in Table 2.2. One trait which is known to influence a species' sensitivity to fragmentation, particularly in birds, is trophic guild. For example, in tropical forests, insectivorous birds have been found to be more sensitive to forest loss than other avian trophic guilds (Şekercioğlu *et al.,* 2002; Bregman *et al.,* 2014). This is because, generally speaking, insectivores have low mobility, are sensitive to small changes in habitat quality (e.g. microclimatic changes), have high levels of habitat specialisation, and are relatively more confined to forest core habitat (Şekercioğlu *et al.,* 2002). The high sensitivity of insectivores to habitat loss is also explained by the food scarcity hypothesis, which postulates that small fragments have reduced invertebrate diversity and thus reduced sources of food for insectivores (Şekercioğlu *et al.,* 2002). The loss of insectivores has important

implications for ecosystem functioning, as insectivores provide a number of different functional roles, such as the control of insect population abundance (Bregman *et al.*, 2014).

Due to the increased recognition of interspecific variations in fragmentation sensitivity, a new type of study has emerged focused on community disassembly (e.g. Ostfeld & LoGiudice, 2003). Just as Diamond (1975b) proposed a number of rules on how ecological communities should assemble, community disassembly studies make predictions on how communities will disassemble following a fragmentation event, and based on species specific traits.

Table 2.2 The traits which dictate species sensitivity to fragmentation. Most studies on extinction risk in fragmented landscapes have been conducted on birds, and thus there is a bird bias in the examples listed in the table. However, many traits are more general and applicable to multiple taxa (e.g. body size). Many of the traits are interlinked, e.g. the relationship between range size and population size (Gaston & Blackburn, 2000).

Trait	Type of species most sensitive to fragmentation	Explanation	References
Body size	With a few exceptions, large bodied species.	Due to high metabolic demand, large bodied species tend to require large home ranges. In contrast, Gotelli & Graves (1990) found that large bodied species were the least extinction prone on landbridge islands, because larger species can a) exploit more resources, and b) due to increased longevity can forego reproduction during times of external stress.	Terborgh (1974); Gotelli & Graves (1990)
Generation time	Long generation time.	Short generation times mean a higher number of emigrants which can colonise and rescue the populations in other patches. They also allow populations to increase their abundance more quickly following disturbances.	Henle *et al.* (2004)
Habitat specialisation	Habitat specialists.	Habitat specialists are acutely dependent on specific resources at all or various stages of their life history, and are thus at risk if these resources are impacted by fragmentation. In contrast, generalists, matrix specialists and edge species benefit from a certain degree of fragmentation.	McCollin (1993); Bommarco *et al.* (2010); Newbold *et al.* (2013, 2014)
Trophic guild	In birds, insectivores are the most sensitive.	Insectivorous birds are generally poor dispersers, sensitive to small changes in habitat quality, have high degrees of behavioural specialisation (e.g. army ant followers), and are generally territorial all year round.	Terborgh (1974); Şekercioğlu *et al.* (2002); Bregman *et al.* (2014)
Range size	Smaller ranged species.	Numerous reasons, including that wide ranging species a) tend to be present in higher abundances in a given site, and b) are predicted to be able to exploit a wider selection of niches.	Koh *et al.* (2004)
Trophic rank	Species higher up the trophic pyramid.	Species at the top of the pyramid have greater metabolic demands and require a larger amount of resources to go through their life history.	Rosenzweig (1995); Henle *et al.* (2004)

Population size in fragment	Species with smaller populations.	Due to environmental, demographic and genetic stochasticity, small populations are generally most at risk of extinction.	Henle *et al.* (2004)
Population size in surrounding landscape	Rarer species, i.e. those with smaller landscape populations.	Patches mirror their surroundings. Thus, if a species is rare in the surrounding landscape, it will generally, be rare in the patch.	Diamond *et al.* (1987)
Population variability	Species with highly variable population sizes.	Species which have highly variable population sizes are more prone to extinction as a) they are often dependent on variable resources which may be lost or affected in the fragmentation process, and b) there is more chance that the population size will drop below a threshold size where stochastic factors (e.g. demographic stochasticity) may push the species over the edge.	Karr (1982); Henle *et al.* (2004)
Dispersal ability	Generally the poorest dispersers, but in some taxa intermediate dispersers may be the most sensitive.	In most taxa, metapopulation theory predicts that poor dispersers are most at risk, as inter-patch dispersal is rare. Thus, small populations are unable to be rescued by immigrants and are more extinction prone. However, in UK butterflies, intermediate dispersers fare worst in fragmented landscapes. This is because they disperse out into the matrix where they suffer high mortality, but are unable to disperse to nearby patches and thus the rescue effect is low.	Hanski (1999); Thomas (2000); Bommarco *et al.* (2010)
Range location	Tropical bird species.	Relative to temperate species, tropical bird species have relatively slow life histories, slow metabolic rates, and low fecundity. Dispersal limitation has also been found to be more extreme in tropical species. Furthermore, tropical species generally have smaller populations. One caveat here is that temperate fragmentation generally occurred centuries ago and thus the most sensitive species may already have been lost.	Robinson (1999); Henle *et al.* (2004); Robinson & Sherry (2012); Salisbury *et al.* (2012)
Endemicity	Regional endemics (here a region may be an island, or a large expanse of forest, e.g. the Amazon).	Generally have restricted distributions, and tend to have quite specialised habitat preferences. Additionally, by definition populations are unable to be rescued by populations outside the region.	Terborgh (1974)

Sociality	Social species, e.g. those which undertake lek mating.	Social species occur in relatively large groups, and thus more habitat and resources are required to maintain the population.	Henle *et al.* (2004)
Colonial nesting (e.g. turtles, birds)	Colonial nesting species.	The eggs of these species are concentrated in certain locations and are thus at risk of being destroyed due to habitat loss, or the various subsequent external threats. A literal representation of the saying "to place all your eggs in one basket".	Terborgh (1974)
Migratory status	Arguments for both migratory and non-migratory species.	Migratory species have been argued to be most sensitive to fragmentation as they suffer from 'double jeopardy.' That is, they are at risk from habitat loss at both ends of their migration route. However, other studies have found that non-migratory species are most at risk.	Terborgh (1974); Newbold *et al.* (2013)

2.1.2.7 Variations in the effects of habitat loss across spatial and temporal scales

Fragmentation is not a linear process, and its associated impacts can vary both spatially and temporally (Haila, 2002; Whittaker *et al.,* 2005; Stiles & Scheiner, 2010). At landscape scales the total amount of habitat in the landscape is likely the most important factor controlling diversity and individual species' incidence (Fahrig, 2013). At the fragment scale, the area of the fragment is the most important driver, and at very local scales (e.g. individual transects), the presence of specific habitat features and density of particular competitor species become more important factors (Stiles & Scheiner, 2010). As a result of this scale sensitivity, the spatial scales of analysis in fragmentation studies can influence the inferences drawn. For example, at local/patch level scales, the decline in habitat specialist and native species may be offset by a rise in generalist and invasive species (Sax & Gaines, 2003), and thus overall richness does not change much. However, at much broader scales (e.g. global) there is a noticeable reduction in the number of species.

Change in assemblage structure due to fragmentation is not an instantaneous process and the relative proportions of each functional species subset, as well as total richness, change throughout time. For example, a chronology of this decline in specialist species is

provided in Fig. 2.4, illustrating the main processes driving this pattern and the sequential order in which each is the dominant process. Again, this is not a linear development, with the proportion of specialists declining at a greater rate to the left of a threshold on a patch area gradient (Humphreys & Kitchener, 1982). In addition, the effects of increased isolation may take many decades to manifest, particularly in plants, as long lived individuals may remain in fragments long after the initial habitat loss occurred (Fig. 2.4; Stiles & Scheiner, 2010).

Analysis of Fig. 2.4 indicates that many species do not go extinct immediately following a fragmentation event; rather, there is a time lag between the event and extinction. MacArthur (1972) theorised that it takes time for a newly created system to reach equilibrium species richness. In the case of land bridge islands, cut off by the rising sea, the resultant island loses species through time and the system relaxes to a new equilibrium (MacArthur, 1972; Wilcox, 1978). Applying this theory to habitat islands, following fragmentation, certain species remain in the fragments consigned to extinction (Fig. 2.4; Tilman *et al.*, 1994; Malanson, 2008). Through time these species become extinct and the habitat island relaxes to a new, lower equilibrium species richness level. These "time delayed but deterministic extinction[s]" have been termed the 'extinction debt' of a system (Tilman *et al.*, 1994, p.65; see also Halley *et al.*, 2013, 2014). Relaxation need not solely relate to extinctions, and can simply refer to the length of time it takes a species' abundance to decline to a lower number following a fragmentation event. In this vein, He & Hubbell (2013) distinguish between two types of relaxation: weak relaxation, which relates to changes in abundance and distribution, and strong relaxation, which relates to species extinctions. Furthermore, an extinction debt need not relate solely to the number of species, and an ecosystem function debt has also been proposed (Gonzalez & Chaneton, 2002).

Relaxation times have been found to vary considerably between taxa and systems (Brook *et al.,* 2003; Halley *et al.,* 2013). In regards to interspecific variation, a number of studies have observed much shorted relaxation times in small bodied taxa with shorter generation times (e.g. invertebrates; Debinski & Holt, 2000; Krauss *et al.,* 2010; Sang *et al.,* 2010). However, even mammal species' assemblages can have short relaxation times. For example, Gibson *et al.* (2013) sampled small mammal species on forest islands in Thailand, created due to the construction of a reservoir. These authors found that 25 years after isolation (i.e. after the reservoir flooded the surrounding land) almost all small mammal species had gone extinct, that is, there was almost complete faunal collapse. In this system, Gibson *et al.* (2013) estimated that the half-life (i.e. the time it takes for a system to lose half its species through relaxation) was only 14 years.

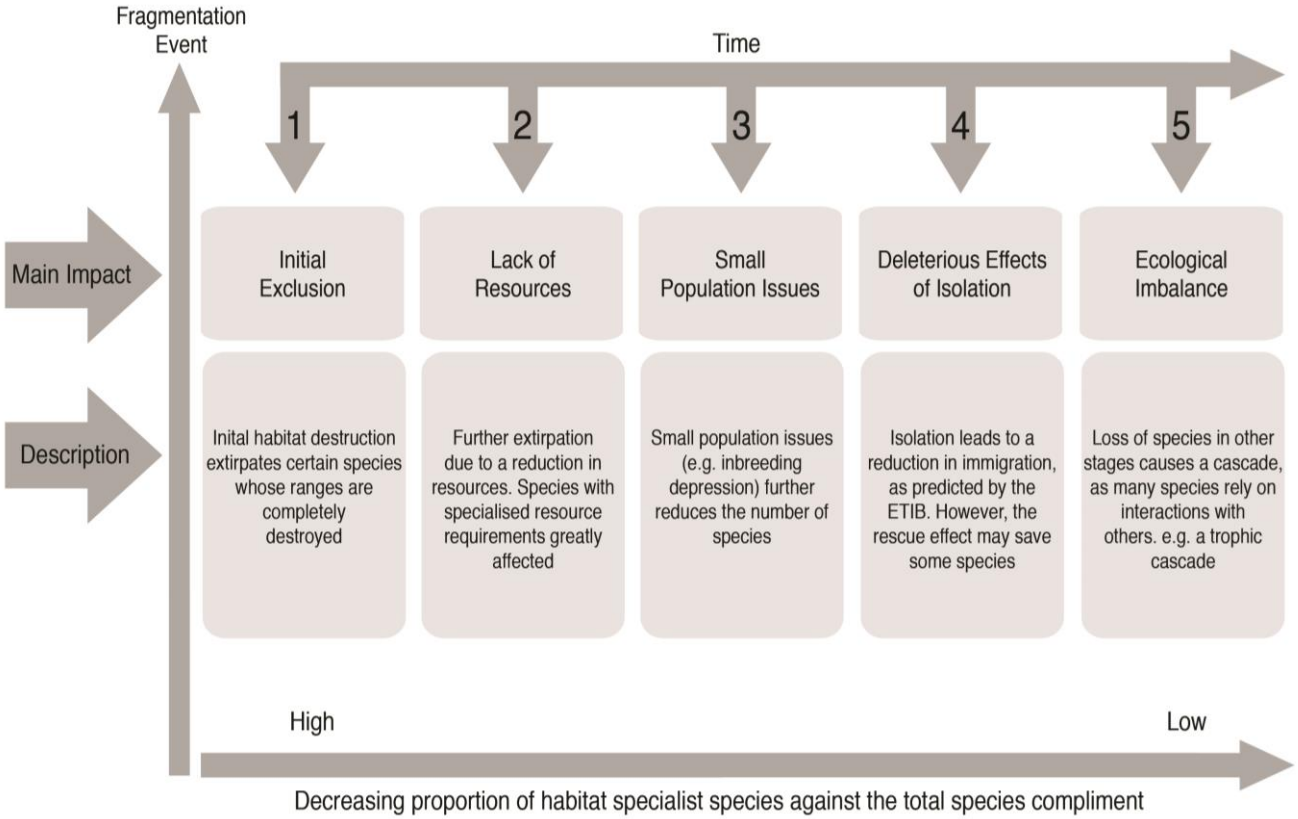


Figure 2.4 Timeline of the decline in species richness following a fragmentation event, listing the main processes facilitating the reduction in proportion of habitat specialists. It should be noted that the increase in habitat generalists can offset the decrease in specialists

and cause total richness to remain static. Based on concepts in Wilcove (1987) and Triantis & Bhagwat (2011).

The relaxation time and extinction debt have been shown to be related to the area of the remaining habitat. For example, it has been estimated that recently isolated 100ha forest fragments have a half-life of less than 15 years (Ferraz *et al.* 2003). In regards to conservation guidelines, Ferraz *et al.* (2003, p.14069) provide the rough rule of thumb that "a 10-fold decrease in the rate of species loss requires a 1,000-fold increase in area." Area is not the only important factor underpinning relaxation times in fragmented systems. For example, nearby fragments may provide immigrants to 'rescue' populations of species in a fragment close to extinction; which will likely delay extinction, although not permanently (Terborgh, 1974).

The presence of this time delayed extinction debt means that predicting extinction as a result of fragmentation is problematic; species inventories are a 'snapshot in time' of this relaxation phenomenon, whereas many predictions, based for example on SAR estimates, have regard to when the 'debt has been paid' (Brooks & Balmford, 1996). The presence of an extinction debt has often been used to explain discrepancies between observed and predicted extinctions due to habitat loss (e.g. Brooks *et al.,* 1999). The time-lag also has positive implications, as it means conservation action can be undertaken to prevent the predicted extinctions from taking place (Brooks *et al.,* 2002). However, as Whittaker *et al.* (2005) attest our knowledge of the power of faunal relaxation is limited, due to the paucity of before and after fragmentation studies (see Laurance *et al.,* 2002). Furthermore, extinction debt and relaxation theory is largely based on MacArthur & Wilson's (1967) equilibrium theory, which, as discussed above, has been criticised in the context of habitat islands (e.g. Laurance, 2008). Therefore, caution is needed when presenting the relaxation phenomenon as a panacea for all studies where predictions do not match empirical findings.

2.1.2.8 One final point on fragmentation

One final point worth considering is that habitat loss and fragmentation rarely occur in isolation from other detrimental changes to natural habitat. In particular, habitat fragmentation is often coupled with a general reduction in the quality of the remaining habitat, i.e. habitat degradation (Fischer & Lindenmayer, 2007). As discussed above, fragmentation increases humans' ease of access into the remaining forest, which brings with it a number of additional activities that reduce the quality of remaining habitat, such as hunting and timber felling for fuel (Laurance *et al.*, 2009). This issue has important conservation implications, as an experimental fragmentation study of butterflies concluded that increasing habitat quality may compensate for a reduction in habitat area, for a number of species (Summerville & Crist, 2001).

2.2 Macroecology and biogeography

> The study of macroecology is not abstracted from the real world. Rather, it helps us understand what causes the world around us to look the way it does.
>
> (Gaston & Blackburn, 2000, p. 300)

2.2.1 Definitions

The overall aim of the thesis is to provide a comprehensive evaluation of macroecological and biogeographical patterns in habitat islands. Thus, it is necessary to define precisely what is meant by the terms macroecology and biogeography. Macroecology can be defined as "a way of studying relationships between organisms and their environment that involves characterizing and explaining statistical patterns of abundance, distribution and diversity" (Blackburn & Gaston, 2002, p. 723). Whereas, biogeography "is the study at all possible scales of analysis of the distribution of life across space, and how, through time, it has changed, within which a major concern has always been the distribution and dynamics of

diversity, frequently codified in terms simply of numbers of species, or proportions of endemic species" (Whittaker, *et al.,* 2005, p. 4). Whilst there are clear differences between the two disciplines (see Blackburn & Gaston, 2002), it is my view that the current thesis sufficiently overlaps both biogeography and macroecology to the extent that it does not make sense to simply refer to it as one or the other. Thus, as stated in the preamble the terms macroecology and biogeography are discussed interchangeably throughout the thesis. The various patterns which form the basis of this thesis are now reviewed.

2.3 The species–area relationship

> Theories, like islands, are often reached by stepping stones. The species–area curves are such stepping stones.
>
> (MacArthur & Wilson, 1967, p. 8)

2.3.1 A brief history

It has often been commented that the species–area relationship (SAR), whereby the number of species increases with area sampled, is the closest thing to a general law in ecology (Newmark *et al.,* 1986; Lawton, 1999; Triantis *et al.,* 2003, 2012; Lomolino *et al.,* 2010a). The SAR has been observed for practically all taxa, and in multiple systems and landscape types (Brooks & Balmford, 1996; Horner-Devine *et al.,* 2004; Magura *et al.,* 2001; Hugueny *et al.,* 2011). The SAR has been documented for over 150 years, with the first species-area curve generally attributed to H.C. Watson in 1859 (Rosenzweig, 1995). However, it was not until work by Arrhenius (1921) and Gleason (1922) that the relationship was expressed mathematically. Arrhenius (1921) noted that the relationship could be explained as a mathematical power function ($S = c\ A^z$, whereby by S and A are the number of species and the area, respectively, and *c* and *z* are fitted constants); although he did not actually record the function. The following year Gleason (1922) criticised Arrhenius' function on the basis it

could not be extrapolated to larger areas, and proposed an exponential function for the relationship. Subsequent papers, such as Preston (1962a,b) have led to an almost universal acceptance of the power function as the superior function to be used in SAR studies; Connor & McCoy (1979), for instance, stating that it has achieved almost paradigm status. Traditionally, the log–log version of the power model ($\log S = \log c + z \log A$) has been employed due to the fact the function is linear in log–log space, enabling easy fitting of the model. However, recent advances in computing power mean the non-linear form can now also be easily fitted.

Although both the power function and the exponential function are the most cited functions in the ecological literature (Tjørve, 2003), they are not universally accepted and since these seminal papers there have been over 20 additional functions published, varying in complexity and the number of parameters involved (Tjørve, 2009; Williams *et al.*, 2009; Triantis *et al.*, 2012), although many functions are just special cases of others (Williams *et al.*, 2009). A more detailed description of the various SAR models is provided in Chapter 5.

There has also been an increasing drive to calibrate the SAR for application in fragmentation studies, incorporating parameters specific to habitat islands, such as matrix permeability and landscape connectivity. A principal criticism with classic SAR models, generally used in combination with a static, arbitrary *z* value (e.g. van Vuuren *et al.*, 2006), are that they ignore the importance of the matrix surrounding the habitat patches (Jules & Shahani, 2003), in particular the permeability and connectivity of the matrix. For instance, Koh & Ghazoul (2010) have shown that by using different *z* values based on the type of the matrix, the predictive power of the power model can be improved by up to 60%. These authors partitioned the standard *z* value into two components: a constant (γ), and a parameter (σ) which characterises the permeability of the habitat matrix for a given taxon and takes a value between zero and one (e.g. $\sigma = 1$ when the matrix is the ocean, and thus in this case the

model reverts back to the standard power model; Koh & Ghazoul, 2010). Other studies have undertaken variations on this approach to include, for example, edge effects and the varying tolerances of different species groups to different habitat types (Pereira & Daily, 2006; Koh *et al.*, 2010; Hanski *et al.*, 2013; Mendenhall *et al.*, 2014a, b), but there is a need for more research in this area, and further broad-scale tests of these previously published 'calibrated' and 'countryside biogeographic' SAR models.

2.3.2 A mélange of curves

While the SAR appears a relatively uncomplicated concept, its application within the ecological literature is somewhat ambiguous, and it is still the topic of critical debate (e.g. He & Hubbell, 2011, and replies by Pereira *et al.*, 2012; He & Hubbell, 2012). Part of this ambiguity concerns the disparity between the different types of SARs. As a first step these different forms can be dichotomised into island species–area relationships ('ISARs') and species accumulation curves ('SACs') (Whittaker & Matthews, 2014; but see Scheiner 2003). The ISAR is a term used to refer to Scheiner's (2003) type IV curve, whereby the number of species occurring within each of a set of islands is analysed as a function of the area of each island (Triantis *et al.*, 2012). The SAC refers to plots of increasing cumulative species number with increasing sampling effort/area (Whittaker & Matthews, 2014). There are multiple types of SACs in recognition of the various ways in which sampling effort can be deployed (see Scheiner's, 2003, type I-III curves). Furthermore, the SAC can be constructed in multiple ways that vary the order in which samples are accumulated to build the curve: in order of increasing or decreasing area, or of increasing or decreasing species richness, etc. (e.g. Quinn & Harrison, 1988; Fischer & Lindenmayer, 2002; Watson *et al.*, 2009; Gavish *et al.*, 2012). Notwithstanding these differences, the concepts are often used interchangeably within the literature. For example, Rosenzweig's (1995) frequently cited three (or four if the point scale is included) component scale SAR typology appears to comprise data sets of two

different structures. Whittaker & Fernández-Palacios (2007) argue that the intra-provincial scale relates to SACs, whilst the archipelagic and inter-provincial scales are ISARs. However, this issue is not articulated by Rosenzweig (1995), nor is it explicitly discussed in many SAR studies.

2.3.3 Mechanisms driving the SAR

There are two main mechanisms theorised to drive the SAR: the area *per se* hypothesis and the habitat diversity hypothesis (Whittaker &Fernández-Palacios, 2007). The area *per se* hypothesis is based on MacArthur & Wilson's (1967) ETIB, and predicates that area intrinsically affects species richness; an area's richness being determined through the equilibrium between immigration and extinction rates. Other authors (e.g. Lack, 1976; Boecklen & Gotelli, 1984; Ricklefs & Bermingham, 2004) argue that factors co-varying with area are more important determinants of an area's richness, rather than a direct area effect, particularly the number of habitats. In this manner the habitat diversity hypothesis stipulates that an increase in patch area equates to greater environmental heterogeneity, for example topographic variation, leading to an increase in the number of habitats and, therefore, niches (Rosenzweig, 1995). For example, Ricklefs & Lovette (1999) studied the SARs of four taxa on the Lesser Antilles, and found that for some, but not all, taxa habitat diversity was more important than area *per se*, and that the role of habitat diversity could be linked to particular biotic traits. A number of additional hypothesis have been proposed to explain the preponderance of SARs in nature, and Table 2.3 provides a relatively comprehensive summary. Whilst most hypotheses are proposed in isolation of other factors, it is likely that many factors act together in tandem to create SARs in real systems. In fact, in most systems it is likely that there is a tapering of effects (Burnham & Anderson, 2002), such that a few factors are the main drivers of the SAR (e.g. habitat diversity), whilst other factors explain smaller amounts of variation (e.g. disturbance).

The SAR is not a scale independent pattern, and the shape of the relationship has been observed and theorised to change with the scale of analysis. For example, in a seminal paper He & Legendre (1996) argued that the exponential model is most appropriate at very small scales, whilst the power model performs better at intermediate scales, and the logistic model is best at very large scales. These changes in best model with spatial scale reflect that the underlying mechanisms also change with scale. At the smallest scales SARs are caused by random sampling effects. At intermediate scales ecological processes drive the relationship, whilst at the largest scales evolutionary and macroecological processes come into play (He & Legendre, 1996; Kisel *et al.* 2011). For example, at the very largest spatial scales the SAR is positive because larger areas contain more biogeographic provinces and thus more unique species (Rosenzweig, 1995).

Table 2.3 The various proposed hypotheses predicting SARs.

Theory	Explanation	References
Equilibrium theory of island biogeography	A larger area can support larger populations, and thus extinction rates are reduced.	MacArthur & Wilson (1967)
Habitat diversity	Increasing area results in greater environmental heterogeneity, for example topographic variation, leading to an increase in the number of habitats and, therefore, niches.	Ricklefs & Lovette (1999); Triantis *et al.* (2003)
Environmental texture hypothesis	Linked to the habitat diversity hypothesis, the environmental texture hypothesis states that the slope of the SAR is related to the rate of change in the environment (i.e. the rate of distance decay in environmental conditions).	Scheiner *et al.* (2011)
Random sampling	Coleman's random sampling hypothesis stipulates that a positive SAR is simply the result of islands randomly sampling individuals from the species pool. Thus, the presence of a species on an island is determined by both the relative area of the island, and the overall abundance of the species.	Coleman (1981); Coleman *et al.* (1982); Bidwell *et al.* (2014)
Minimum area	Many species have minimum area requirements,	Diamond

requirements/incidence functions	i.e. a patch size below which they are unable to go through all stages of their life history. If these minimum area requirements vary across species, then summing individual species incidence functions results in a positive SAR.	(1976); Hanski (2010)
Fractal distributions	Self-similar distributions of individuals and species across space have been shown to result in power law SARs. It has also been shown that a multifractal approach can generate triphasic SARs.	Harte *et al.* (1999); Borda-de-Agua *et al.* (2002)
Neutral theory	A variety of neutral models have been found to produce realistic looking SARs.	Bell (2001); Rosindell & Harmon (2013)
Energy–area relationship	The increase in species with area is actually driven by an increase in the amount of available energy with increasing area.	Wright (1983); Jonsson *et al.* (2011)
Speciation–area relationship	Recorded speciation events increase with area, and above a certain threshold it is speciation which drives the SAR. There are many reasons why speciation rates might increase with area, including an increasing opportunity for geographic isolation and allopatric speciation, and an increase in the number of habitats/niches. This hypothesis is more relevant to oceanic islands than habitat islands.	Losos & Schluter (2000); Wagner *et al.* (2014)
Core:edge ratio	Larger areas have more undisturbed core habitat, which in turn supports more patch specialist species.	Fischer & Lindenmayer (2007)
Disturbance hypothesis	Smaller islands suffer more disturbance events, and the effect of disturbance, both natural and anthropogenic, is amplified in smaller islands.	Triantis *et al.* (2006)

2.3.4 Biological interpretation

Early SAR studies treated the parameters of the various models as non-biological constants (see Connor & McCoy, 1979). However, following MacArthur & Wilson and Preston's work on the SAR, numerous studies have attempted to determine whether SAR model parameters can be interpreted biologically. As almost all of this work has focused on the parameters of the power log–log model (i.e. *c* and *z*), this section does likewise. This debate has often become heated, with some authors vehemently arguing that *c* and *z* can be interpreted

biologically (Rosenzweig, 1995), whilst others countering with the position that they are simply mathematical parameters with no biological meaning (e.g. Connor & McCoy, 1979; Abbott, 1983).

Preston's canonical lognormal SAD theory (discussed in Section 2.7) predicted *z* values of roughly 0.25. However, whilst some studies have reported *z* values in the region of 0.25, a wide range of *z* values have been observed. For example, in a study of over 600 true island datasets, Triantis *et al.* (2012) found the mean *z* value to be 0.32, but the range stretched from 0.06 to 1.31. Part of this variation is due to the fact that *z* has been found to vary with a wide range of system properties. For instance, due to a more permeable matrix, *z* values have been predicted to be lower in habitat islands relative to true islands (Watling & Donnelly, 2006). In addition *z* has been found and proposed to: i) increase for taxa with poorer immigration abilities (Lomolino, 1982; but see Aranda *et al.*, 2013), ii) increase for taxa with higher food plant specialisation and trophic rank (Holt *et al.,* 1999; Steffan-Dewenter & Tscharntke, 2000), iii) increase for taxa with a larger average body size (Azovsky, 2002), and iv) vary with many other taxon specific characteristics, such as energetic demands and habitat requirements (Marquet *et al.*, 2004).

The *c* parameter, sometimes referred to as the 'forgotten parameter', has been much less well studied than *z* but has been proposed to reflect a number of properties, including the average size of the most common species, a measure of carrying capacity, and a scale-independent measure of diversity (Gould, 1979; Whittaker & Fernández-Palacios, 2007; Triantis *et al.,* 2012). Rosenzweig *et al.* (2011) have proposed that *c* defines species density and is thus useful for comparing density across samples with different SARs. As with *z, c* has been found to vary with a number of factors, including latitude, across different taxa, the degree of isolation and the scale at which area is measured (Gould, 1979; Connor & McCoy, 1979; Wright, 1983; Sólymos & Lele, 2012). In reality, *c* and *z* are not independent and are

negatively correlated; thus, studies should always bear this in mind when trying to attach any meaning to either (Gould, 1979).

2.3.5 Applied uses of the SAR

> The SAR is a fundamentally important pattern within ecology and understanding the processes that shape the particular form taken by each type of SAR is crucial for both ecological biogeography and conservation biogeography.
>
> (Whittaker & Triantis, 2012, p. 623).

The SAR is a simple but powerful tool with numerous applied applications. For example, constructing baselines and targets for conservation strategies (Desmet & Cowling, 2004), upscaling diversity in areas too large to sample (Smith, 2010), identifying biodiversity hotspots (Guilhaumon *et al.,* 2008), and guiding protected area network design (Diamond, 1975a; Whittaker & Fernández-Palacios, 2007). However, perhaps the SAR's most prevalent application is in the prediction of extinctions resulting from habitat loss (He & Hubbell, 2011; Halley *et al.,* 2013). The common procedure in such studies is to apply the power function backwards, i.e. to calculate the number of species lost due to a reduction in habitat area (e.g. Brooks & Balmford, 1996; Brook *et al.,* 2003; Koh & Ghazoul, 2010; Whittaker & Matthews, 2014). Generally a set *z* value of around 0.18-0.25 is used. Although widely used as a predictive measure, it has been noted that predicted extinctions commonly exceed observed extinctions (Brooks *et al.,* 1999; Halley *et al.,* 2013). This discrepancy has been attributed to various factors, including the possibility of an extinction debt (Tilman *et al.,* 1994), the fact that as habitat is lost the parameters of the model change but predictions are based on the initial parameter values (Smith, 2010), and the observation that, in contrast to oceanic islands, the habitat matrix surrounding fragments can act to ameliorate the impacts of a reduction in primary habitat extent (Ricketts, 2001; Baum *et al.,* 2004). In addition to area,

other aspects of fragmented landscapes likely affect extinction, and are not incorporated in simple SAR calculations, e.g. hunting pressure (Benchimol & Peres, 2013).

It is also possible that in many systems the conservation actions of humans have prevented a number of species from becoming extinct that otherwise would have based on the reduction in habitat amount. Due to the possibility of an extinction debt and the positive impact of conservation actions, a number of studies have used the number of species classified as endangered and critically endangered by the IUCN as the variable of interest rather than the number of observed extinctions; and have subsequently observed closer matches between predicted and observed "extinctions" (see Halley *et al.*, 2013).

A further problem with the SAR is that by definition it incorporates all species in an area. However, the only species which can possibly be driven to extinction in the area due to habitat loss are those species endemic to the area. Thus, a number of studies have recommended the use of the endemics–area relationship rather the standard SAR, when predicting extinctions (Smith, 2010; He & Hubbell, 2011; but see Matias *et al.*, 2014).

Other problems with using the SAR to predict extinctions are that it is not possible to calculate an associated measure of uncertainty, and that when using the SAR in such a way, one implicitly assumes that a species is present in the landscape if one individual remains; a fact that is obviously true in the literal sense, but which is evidently problematic for conservation purposes (Kitzes & Harte, 2014). To address these shortcomings, Kitzes & Harte (2014) have recently introduced two new metrics, termed the extinction–area relationship (the probability of a given species going extinct following habitat loss) and the probabilistic species–area relationship (the probability a species is protected in the landscape following habitat loss). These metrics are based on empirically sound SADs and spatial abundance distributions, and have been found to provide more accurate predictions than the standard SAR approach.

Despite these criticisms, a strong position can be taken that the SAR is a valuable conservation and policy 'tool' as it simple to use, and has the potential to make semi-reasonable predictions without the need for large amounts of species-specific data. Accordingly, research should be undertaken to calibrate it with the factors distinctive to terrestrial systems (e.g. connectivity) so that it can be more effectively applied to terrestrial fragments rather than discarded as a research approach, as certain authors have argued (Laurance, 2008).

2.4 Habitat amount thresholds

2.4.1 Types of threshold

A recurring question over the last two decades has been whether there exists a threshold in the relationship between habitat amount and species richness; such that below a threshold value there is a precipitous decline in species richness. In this context, thresholds can be defined as "points or zones of abrupt change in ecological relationships" (Huggett 2005, Groffman *et al.*, 2006; Ficetola & Denoël, 2009); that is, an additional unit change in a particular variable (e.g. amount of habitat) at the break-point causes a rapid shift in ecological conditions. The concept of thresholds in ecology and conservation is broadly employed and used to examine numerous relationships. For instance, in the study of the minimum viable area for population persistence (Shaffer, 1981), biological control of pests in agro-ecosystems (With *et al.*, 2002), percolation theory and metapopulation extinction (Bascompte & Sole, 1996), Allee-like thresholds in fragmented systems (Amarasekare, 1998), and the existence of critical threshold levels of habitat loss and fragmentation that result in sudden reductions in species' occupancy and overall richness (With & Crist, 1995).

There are two types of habitat amount threshold studies: (1) landscape habitat amount thresholds (e.g. Fahirg, 2001), and (2) thresholds in the ISAR (e.g. Drinnan, 2005). Most

research has focused on the first type of threshold, and studies of this nature are interested in the amount of natural habitat in entire landscapes. The second type of threshold has been less studied, perhaps because thresholds are harder to test for in non-linear functions. Most SAR models are non-linear (e.g. the power model), and whilst some of the functions can be linearised through logarithmic transformations (e.g. the power model; Triantis *et al.*, 2012); the problem then becomes how to interpret a threshold in logarithmic space. The increasing uptake of piecewise regression models (i.e. models that fit two separate linear functions to upper and lower parts of the distribution of values of area; Gentile & Argano, 2005) represents a promising method with which to test for thresholds, but still the question of logarithmic transformation remains (Burns *et al.*, 2009).

2.4.2 Evidence for habitat amount thresholds

2.4.2.1 Landscape thresholds

Ascertaining whether landscape thresholds are a natural phenomenon, along with efforts to estimate threshold values for guiding land management policy and protected area network design has been a major focus of conservation science in the past twenty years in an attempt to aid decision makers in managing trade-offs between anthropogenic land uses and conservation land uses. The principal thrust of this research has centred on theoretical modelling (e.g. With & Crist, 1995; Fahrig, 1997, 2001, 2002; With & King, 1999a) and these studies have produced wide-ranging results. A seminal review of landscape threshold studies by Andrén (1994) concluded that a significant ecological extinction threshold is crossed when landscape habitat cover is reduced to 10-30%; this being the point where the detrimental effects of fragmentation act to intensify the effects of habitat loss, resulting in the loss of species. Other theoretical studies have stated that threshold values vary between taxa and regions, with published values ranging from 1% to 99% of landscape habitat cover (With & King, 1999a; Fahrig, 2001; Zuckerberg & Porter, 2010). Despite the predominance of

simulation models in the literature, a rising number of empirical studies have investigated the existence of landscape thresholds, with many offering data on the existence of a threshold (Radford *et al.,* 2005; Zuckerberg & Porter, 2010; see Huggett, 2005 for a review). Yet, several studies found no evidence of landscape thresholds. For example, in their study of potential threshold effects of bird and reptile assemblage responses to habitat loss in a plantation landscape in New South Wales, Australia, Lindenmayer *et al.* (2005) failed to find any evidence of threshold relationships between native habitat cover and bird and reptile diversity, despite using appropriate statistical techniques.

2.4.2.2 Drivers of landscape habitat amount thresholds

If the significance of time since fragmentation and any associated extinction debt (Tilman *et al.,* 1994; With & King, 1999b) are momentarily discounted, the causes of landscape thresholds can be summarised into two main hypotheses: the extinction threshold hypothesis and the fragmentation threshold hypothesis. The extinction threshold hypothesis postulates that the effect of habitat loss becomes exacerbated at low levels of habitat cover resulting in the average patch size becoming very small, and in turn resulting in the patch mortality rate exceeding its reproductive rate (Fahrig, 2001, 2002). Below such levels there are not enough resources within the patch to maintain populations, resulting in numerous species being extirpated from the patch. At these low levels of habitat cover, factors such as edge effects also act to reduce species richness at an accelerated rate (Drinnan, 2005). As the number of individuals decline, the remaining populations contain fewer individuals leading to the oft-reported environmental, demographic and genetic stochasticity problems that increase the risk of local extinctions of small populations (Lindenmayer & Luck, 2005; Primack, 2010). This reduction in population size beyond a threshold amount of habitat is also a likely primary cause of thresholds in habitat island ISARs.

The fragmentation threshold hypothesis on the other hand infers that thresholds are driven by an increase in the effects of fragmentation below the threshold habitat amount; that is, where the effects of fragmentation intensify the deleterious effects of habitat loss, leading to a rapid reduction in species richness in patches below the threshold (Andrén, 1994; Zuckerberg & Porter, 2010). Yet, there has only been moderate empirical support for this theory (Fahrig, 2003; Betts *et al.,* 2007; Pardini *et al.,* 2010), and certain studies (e.g. Gavish *et al.,* 2012) have reported that patch species richness actually benefits from a degree of fragmentation, potentially the result of higher matrix survival rates (Fahrig, 2002). However, certain studies have stated the importance of isolation and its role in determining dispersal thresholds. For instance, With & King (1999a) examined percolation and dispersal success thresholds in fragmented landscapes through simulation models and found a 'critical gap-size structure' after which inter-patch dispersal is disturbed.

2.4.2.3 ISAR thresholds

The search for thresholds/breakpoints in the ISAR of habitat islands has occurred at a much slower pace than the search for landscape thresholds. In fact, almost all the work on ISAR thresholds has been focused on oceanic islands, under the guise of the small island effect (SIE). The SIE can be defined as the situation where "species number does not increase with increasing area in small islands. In short, below a certain threshold area diversity is independent of island size." (Triantis *et al.,* 2006, p. 915). Although mentioned previously, the SIE concept was popularised by the work of Lomolino and colleagues (e.g. Lomolino, 2000, 2002; Lomolino & Weiser, 2001). Lomolino argued that the ISAR is not a monotonic relationship; rather, it is comprised of three components and the shape is in fact sigmoidal in arithmetic space. According to this argument, at small areas richness varies idiosyncratically with area (the SIE), and at larger areas the relationship approaches an asymptote. However, this theory has been vigorously debated (see Williamson *et al.,* 2001; Lomolino, 2002).

Furthermore, tests of this sigmoidal model using multiple datasets have indicated it rarely provides the best fit to empirical data, possibly because the range of island areas in these studies are far below what Lomolino had in mind (e.g. Triantis *et al.,* 2012). In fact, tests for the SIE in general have produced mixed results (Triantis *et al.,* 2006; Dengler, 2010; Triantis & Sfenthourakis, 2012).

One exception to the above statement that most work on ISAR thresholds has focused on oceanic islands is the study of Drinnan (2005). Using a small selection of ISAR models, Drinnan (2005) looked for thresholds in the ISAR of multiple taxa in a set of urban fragments in Sydeny, Australia. Patch area–richness thresholds were observed at 2ha for plants and fungi, and 4ha for birds and frogs.

2.5 Nestedness

While the idea was discussed earlier (e.g. Darlington, 1957), nested structure was first formally defined in the 1980s as the situation in which depauperate island faunas constitute proper subsets of the species in more species rich island faunas (Patterson & Atmar, 1986). This definition was based on the idea of ordering a presence/absence matrix of species on islands in relation to island species richness. However, subsequent authors have used other variables (e.g. isolation) to order the presence/absence matrix (e.g. Kadmon, 1995; Wang *et al.,* 2010). Hence, nestedness may now be regarded simply as a form of ordered composition of species assemblages involving a significant tendency for packing of the matrix into a series of proper subsets, with a variety of mechanisms identified as potentially contributing to patterns of nestedness or anti-nestedness differing from random expectation (discussed in depth in Chapter 8). Anti-nestedness is a term encompassing several patterns of community structure: high turnover (Leibold & Mikkelson, 2002), perfect checkerboards (Diamond,

1975b), or the deviation from nestedness whereby within a set of sites, species are only present in a single site (Poulin & Guégan, 2000).

A number of nestedness metrics have been proposed and debated (Atmar & Patterson, 1993; Cutler, 1994; Wright *et al.*, 1998; Almeida-Neto *et al.*, 2008; Baselga, 2012). Early metrics such as N0 (Patterson & Atmar, 1986) were based on the approach of counting the number of unexpected gaps in matrices; a method which has since been found to be dependent on matrix size, shape and fill, and which produces inconsistent results (Brualdi & Sanderson, 1999). Although originally popular, the matrix/nestedness temperature (Atmar & Patterson, 1993) and discrepancy metrics (Brualdi & Sanderson, 1999) have also been criticized for dependence on matrix size and shape, and for measuring something different from the nestedness concept as defined by the metrics' originators (Ulrich & Gotelli, 2007a; Almeida-Neto *et al.*, 2007, 2008). The recently proposed NODF metric (Almeida-Neto *et al.*, 2008) appears to circumvent many of the aforementioned issues and represents a promising tool. Alongside the problem of determining an appropriate metric of nestedness, there has been debate about how to determine the significance of observed departures from random expectations. The usual protocol has been to compare the observed nestedness metric value with a distribution of values generated using null communities, with much debate ensuing about the choice of algorithms for simulating the null communities (Wright *et al.*, 1998; Miklós & Podani, 2004; Ulrich & Gotelli, 2007a; Almeida-Neto *et al.*, 2008).

Early studies found significant nestedness to be near universal (e.g. Simberloff & Martin, 1991; Atmar & Patterson, 1993) and it was thus generally assumed that nestedness was a common pattern in island systems (Wright *et al.*, 1998; Whittaker & Fernández-Palacios, 2007). However, the majority of these early studies used metrics and null community algorithms based on the concept of random fill, an approach which has since been criticized as inappropriate (Ulrich & Gotelli, 2007a, b). In particular, it has been argued that

many of the null community algorithms used are biased as they do not keep matrix row and column totals constant (Gotelli & Entsminger, 2001a; Miklós & Podani, 2004). The application of more appropriate algorithms, such as the 'swap' algorithm and its variants, and metrics such as NODF, has led workers to question whether significant nestedness is as common as once assumed (e.g. Ulrich & Gotelli, 2007a; Almeida-Neto *et al.*, 2008).

Nestedness is important from a conservation biogeographic perspective because it potentially informs protected area placement and design in fragmented and degraded landscapes (Cutler, 1994; Fischer & Lindenmayer, 2005a, b; Triantis & Bhagwat, 2011; but see Simberloff & Martin, 1991), contributing to the SLOSS ('single large or several small') protected area debate (see Ovaskainen, 2002). For example, if a set of habitat fragments in a landscape are perfectly nested (*sensu* Patterson & Atmar, 1986) all the species in the fragments can be represented at least once merely by conserving the most species rich fragment. However, due to differences in the way in which species respond to fragmentation (see section 2.1), perfect nestedness is extremely rare, and thus the situation is rarely so simple (Fischer & Lindenmayer, 2005b). Moreover, the search for general nestedness patterns in habitat islands is hampered by the fact that studies have found the degree of nestedness to vary according to a number of factors, including taxon (Fischer & Lindenmayer, 2005a) and the type of matrix (Wethered & Lawes, 2005). Due to these observations, Fischer & Lindenmayer (2005a) have argued that nestedness theory should only be applied in conservation contexts alongside detailed autecological studies of the relevant species.

Traditionally, nestedness studies have focused solely on taxonomic nestedness, i.e. nestedness involving species richness. However, recently nestedness analyses have been extended to incorporate differences between species in terms of functional traits (functional nestedness) and evolutionary history (phylogenetic nestedness) (Melo *et al.*, 2014). These

new indices (e.g. treeNODF; Melo *et al.*, 2014) measure nestedness using the presence-absence matrix in combination with a tree like structure (e.g. a phylogeny or a trait dendrogram), and can be used to test for nestedness whilst taking into consideration how similar species are in terms of their traits or evolutionary history. This approach is potentially revealing, as a system can be functionally and/or phylogenetically nested, but not taxonomically nested. These indices are very recent developments, and no studies have examined patterns of functional and phylogenetic nestedness using multiple datasets.

2.6 Metapopulations

A single species metapopulation is a spatially structured set of subpopulations (patches) which are connected through interpatch dispersal (Levins, 1969; Hanski, 1997, 1998, 1999). The primary processes in a metapopulation are the extinction of subpopulations due to various factors, such as low habitat quality or environmental stochasticity, and the colonisation of vacant patches (Hanski, 1999). Thus, a characteristic of most metapopulations is the turnover of populations in individual patches, whilst the metapopulation as a whole survives. In real world fragmented landscapes, variation in patch size and isolation will underpin the extinction and colonisation rate among patches (Opdam, 1991; Hanski, 1994a). Precisely how the extinction and colonisation rates are linked to patch variables depends on the metapopulation model employed. In Levins' (1969) original metapopulation model, the extinction rate depends only on P, the proportion of occupied patches (i.e. the proportion populations that can become extinct), and the colonisation rate depends on both the number of occupied and unoccupied patches. Hanski (1982) advanced Levins' model by assuming the extinction rate to be a quadratic function of P. In Hanski's model, the extinction rate rises at first with P, but then decreases as more patches become occupied. This decrease in extinction rate at high values of P represents the rescue effect, in which populations are saved from

extinction by immigrations from neighbouring patches (Hanski, 1982). Subsequent authors have made further adaptations to these metapopulation models (see Gotelli, 1991; Hanski, 1999; Gaston & Blackburn, 2000). For example, Gotelli (1991) introduced a model which assumes that immigration is not driven by local populations, but from a constant propagule rain (i.e. the colonisation rate is independent of P). There are also different variants of the metapopulation concept, with the two main variants being the classic metapopulation in which the individual patches are of similar size, and the source-sink variant in which one patch, often termed the 'mainland' or 'continuous habitat,' is much larger than the other patches in the network and acts as a source of immigrants (Whittaker & Fernández-Palacios, 2007). This is a slight simplification as the source patch is not necessarily always the largest in the network; for example, the source could be a small patch with high quality habitat (Pulliam, 1988).

Whilst metapopulation theory is well suited to certain taxa, e.g. butterflies (Hanski *et al.,* 1996), most species likely do not follow strict metapopulation dynamics (Gotelli, 1991). However, as Ovaskainen (2002) attests, this does not stop the theory providing a useful lens for viewing fragmented landscapes through.

2.7 Incidence functions

2.7.1 A brief history

Incidence function models predict spatial patterns of occupancy for single species, as a function of particular site or landscape characteristics (MacKenzie *et al.,* 2006; Rueda *et al.,* 2013). The incidence function concept was introduced by Jared Diamond in his work on the distribution of birds in New Guinea (e.g. Diamond, 1975b, 1976). In his early work Diamond (1975b, p353) argued that incidence functions "[serve] as a fingerprint of the distributional strategy of a species," going on to state that species distributions "define a spectrum of

incidence functions," classified into six categories. At one end of the spectrum Diamond (1975b) characterised the so called 'super tramp' species. These species represent the extreme of the *r* selection strategy with a high tolerance of environmental stress and no specific habitat specialisation. The super tramps' incidence functions are characterised by a non-zero incidence on species poor islands, while being entirely absent on species rich islands. At the other end of the spectrum are 'High S' species, representing the reverse of super tramps. For example, High S species typify the extreme of the *k* selection strategy and are confined to the most species rich islands (see Diamond, 1975b, for an explanation of each category). It should be noted that much of Diamond's work on incidence functions and assembly structure has been widely contested (e.g. Connor & Simberloff, 1979).

The incidence function concept underwent a revival during the development of metapopulation theory. Ilkka Hanski (e.g. Hanski, 1992, 1994a, 1997, 2010) took Diamond's incidence function model and adapted it for modelling metapopulation dynamics. Hanski's metapopulation models incorporate both area and isolation, and can be used to calculate the probability of occupancy, and the fraction of patches in a metapopulation predicted to be occupied, through defining an explicit function linking extinction and colonisation rates to patch area and isolation. Thus, Hanski's models are spatially realistic in the sense that they incorporate actual information on patch characteristics. Various metapopulation incidence function models have been proposed of differing complexity; for example, depending on whether a mainland patch is present or absent (Hanski, 1994a). One of the benefits of these models is that metapopulation dynamics are assumed to represent a stationary Markov process (i.e. each patch has a constant transition probability between being occupied or not occupied by a given species), and thus parameters characterising extinction and colonisation rates can be estimated from snap shot presence-absence data, which avoids the complicated task of attempting to empirically measure extinction and colonisation rates (Hanski, 1997).

For example, one of the most basic of Hanski's incidence function models assumes that if a patch (*p*) is occupied, the population will go extinct with probability (E*p*) in one time unit (Hanski, 1997). If *p* is empty, it is colonised with probability C*p* in one time unit. The probability that *p* is occupied can then be calculated by:

$Jp = Cp/(Cp + Ep)$,

E*p* is a function of island area, such that:

$Ep=\min[\mu/A^{x}, 1]$,

whereby A is the area of *p*, and x and μ are parameters generally estimated from the data. Calculating C*p* depends on the type of model being used (e.g. whether or not a mainland is included or not), but if a mainland is assumed one derivation is:

$Cp = \beta e^{-\alpha di}$,

whereby di represents the distance between *p* and the mainland, and α and β are parameters. These four parameters (x, μ, α and β) are generally estimated using maximum likelihood methods, but in certain cases particular parameters have been empirically measured (e.g. Hanski, 1994b). These metapopulation models are interesting but are generally only of use when a species is distributed across a landscape as a functioning metapopulation, which has not been proven for many taxa (Fronhofer *et al.,* 2012).

More recently, additional incidence function models have developed which account for imperfect detectability of species and missing data (e.g. MacKenzie *et al.,* 2002; Risk *et al.,* 2011). False absences are argued to be particularly problematic for incidence function models (MacKenzie *et al.,* 2006). However, many of these models have been developed in a Bayesian context (e.g. Royle & Kéry, 2007; Risk *et al.,* 2011) and require presence-absence data for a set of patches across multiple years.

2.7.2 The independent variable in incidence function models

Numerous factors have been hypothesised to determine species' patch occupancy probabilities. For instance, as stated above, the habitat diversity hypothesis postulates that the number of habitats increases with patch/island area (Ricklefs & Bermingham, 2004). Thus, species with high levels of habitat specialisation are likely to have zero-incidence on smaller patches with low habitat diversity. Conversely, habitat generalists are likely to exhibit the opposite incidence function, having a high incidence on disturbed, smaller patches, but being out competed in larger patches with high numbers of habitat specialists (Diamond, 1975b; Diamond, 1979). Perhaps the most widely cited factor determining a species' incidence is the influence of area (see Whittaker & Fernández-Palacios, 2007, and references therein). Once the co-variation with habitat diversity has been discussed, area can be summarised as having three main influences on incidence. Firstly, a species will only be present in a patch which fulfils its resource requirements; although, it is important to remember a species can potentially acquire its resources from multiple patches if it is able to disperse across the matrix (Diamond, 1975b). Secondly, species will only be continually present on islands which are large enough to sustain the minimum viable population of the species (Gilpin & Soulé, 1986; Primack, 2010). Finally, and linked to the issue of minimum viable populations, MacArthur and Wilson dynamics imply that larger areas have lower extinction rates (MacArthur & Wilson, 1967). Thus, the ecology of individual species (e.g. reproductive ability, competitive ability) will determine the species' area sensitivity in response to extinction and immigration dynamics. Whilst acknowledging that other variables can be used to construct incidence functions, from herein I use the term incidence function solely to refer to models constructed using island area as the independent variable.

2.7.3 Fitting standard incidence function (area) models

Standard incidence function models (i.e. those which are the focus of my Chapter 12 analyses) are relatively straight forward to fit to empirical data. Generally they involve fitting a generalised linear model (GLM) from the binomial family, whereby the presence-absence of a species in a set of patches is used as the response variable, and the area of each patch is used as the predictor variable (Watson *et al.,* 2005; MacKenzie *et al.,* 2006). Although not a universal observation, when constructed using a binomial GLM, the incidence function of most species is characterised by a squashed S-shaped curve (e.g. Fig. 2.5). Thus, the steepest change in the probability of occurrence occurs in the lower-middle part of the area gradient. Whilst the response variable represents a binomial outcome (i.e. a species is either present or absent), the binomial GLM returns a probability, which can be interpreted as the probability that a species is present on an island of a given size. The general protocol employed when using incidence function models in conservation focused studies has been to select a specific probability of presence value as the value of interest. This means selecting the island area value which corresponds to a given probability of occurrence. Various values have been used ranging from 30% to 60% (van Dorp & Opdam, 1987; McCoy & Mushinsky, 2007; Rueda *et al.,* 2013; Pe'er *et al.,* 2014). However, perhaps the most commonly used probability value, and the value used in the current thesis, is 50% (0.5, see Fig. 2.5; e.g. Watson *et al.,* 2005; Hanski, 2010). Hanski (2010) refers to the 50% probability level as the 'critical island area', and from herein when I refer to comparing incidence functions across landscapes I mean comparing critical island area values (CIVs). For example, comparing the incidence function of species X based on data from landscape A and landscape B, would involve calculating the critical island area for both A and B and determining which is lower/higher.

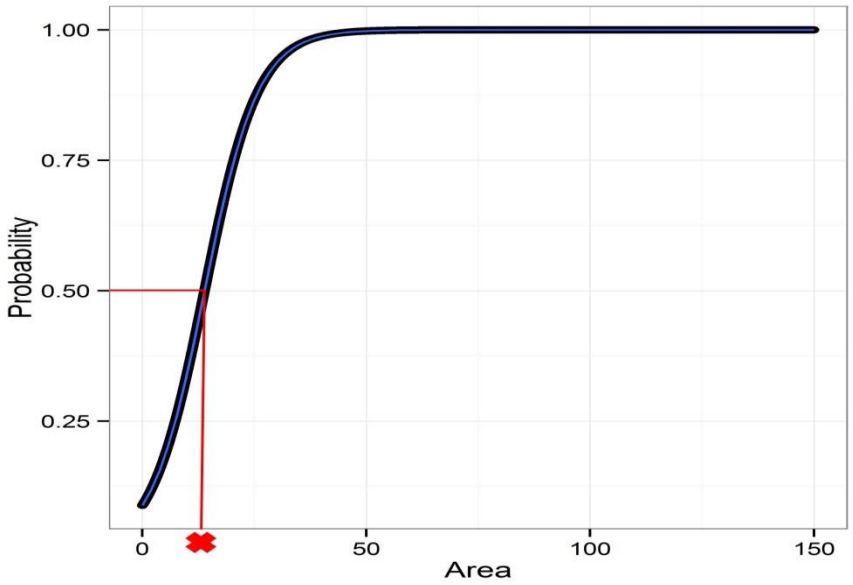


Figure 2.5 An example incidence function plot constructed using simulated data. The critical island area (i.e. the area corresponding to the 50% incidence probability) is marked with a red cross.

2.7.4 Applied uses of incidence function models, and associated research gaps

Incidence function models have been argued to represent useful conservation tools, and can be used in a number of conservation applications, including protected area design, endangered species action plans, and assessing the extinction risk of individual species (Hanski, 1994a, 1997; Araujo & Williams, 2000; Hu *et al.*, 2012). The effects of fragmentation are known to vary across species (McCollin, 1993) and incidence functions are a useful method for teasing out these differences (van Dorp & Opdam, 1987; Opdam, 1991). In this way, the use of incidence function models is an answer to those who have argued for a greater focus on the 'biogeography of the species' (e.g. Opdam, 1991; Lomolino *et al.*, 2010b) and the focal species approach (Lambeck, 1997). The focal species approach is akin to the umbrella species concept, and aims to select particular 'focal species' based on set criteria, on the basis that conserving this small set of focal species will in turn result in the conservation of the majority of species (Lambeck, 1997). For example, Nicholson *et al.* (2013) showed that a focal species approach resulted in a successfully designed reserve system for a larger number of vertebrate species in Australia. Various criteria have been proposed for selecting focal species, including using incidence function type models to

calculate critical island areas and minimum area requirements for multiple species (Watson *et al.,* 2001; Pe'er *et al.,* 2014).

As the above examples attest, in theory incidence functions represent useful conservation tools. However, the real utility of incidence functions rests on the assumption that a given species' incidence function does not vary to any substantial extent across space and time (MacKenzie *et al.,* 2006). For example, it would be no use designing a protected area network in the location B based on the results of an incidence function study of a species in a given location (A) and year (X), if it was found that the incidence function of the same species was substantially different in location B or year Y. A number of studies have attempted to test this stability assumption, at least implicitly, with generally negative results. For example, Watson *et al.* (2005) compared the incidence functions of bird species in forest fragments distributed across three matrix types: urban, peri-urban and agricultural. Their study found that CIVs varied considerably for individual species across the three matrices. To take another example, Hinsley *et al.* (1994, 1996) observed that the incidence functions of certain bird species in UK forest fragments varied across years due partly to changes in weather/climatic conditions across years (e.g. harsh winters). One proposed work around to deal with the issue of varying incidence function model parameters across years is to measure many independent patch networks across space (ter Braak *et al.,* 1997). However, this only works if there is not substantial variation in incidence functions across space. The results of Watson *et al.* (2005) indicate that in such an endeavour, matrix type would need to be kept constant at the very least. But do incidence functions vary spatially in other ways? One largely unexplored research avenue in this regard is whether incidence functions vary in a systematic manner across a species' range. This possibility resulted in the generation of a testable hypothesis which formed the basis of my field work data collection and Chapter 12 analyses. The null and alternative hypotheses were:

H0 – For a given species, there is no systematic variation in the critical island area value between landscapes located in the range core and the range edge.

H1 – For a given species, the critical island area value will be lower (i.e. a smaller area) for landscapes located in the range core than those located at the range edge.

The hypothesis that the CIV will be lower at the centre of the range relative to the periphery is based on the core-abundance hypothesis. Despite the generalisation that abundant species are consistently abundant, and rare species are universally rare, species are not continuously or evenly distributed across their range. Rather their distribution can be described as patchy with large gaps where the species does not occur. Similarly, the abundance patterns of most species are represented by numerous sites of relatively low abundance, and a relatively few sites of considerably high abundance (Gaston, 1994; Brown *et al.,* 1995; Gaston & Rodrigues, 2003). Even species that are regionally abundant may have reduced abundances in certain areas across the range, in accordance with 'Taylor's power law' (Taylor, 1961). This pattern is not static in space or time, however, and the peaks of abundance have been found to move spatially between generations; this being a process not generally well understood (Taylor, 1986; Gaston, 1990).

The core-abundance hypothesis postulates that species have higher incidence and are more abundant in their range core, and that abundance decreases towards the range margins. The core-abundance hypothesis forms a central component of numerous ecological hypotheses, for example directions of gene flow within species and the spatial distribution of pest outbreaks, and thus the pattern is generally regarded as fact by many researchers. In his seminal work on plant distributions along environmental gradients Whittaker (1967; also see discussion in Brown, 1995) used the normal distribution to highlight how abundance peaked in the range core and decreased towards the edge of the range. A review by Hengeveld &

Haeck (1982) concluded that generally a species' abundance is normally distributed across its range, albeit with an amount of scatter and that this pattern is present in numerous taxa and is independent of spatial scale.

Numerous explanations have been put forward in an attempt to explain the core-abundance pattern. An early descriptive explanation of the pattern likened dispersing organisms to gas molecules dissipating from a central nucleus of high abundance to areas of lower abundance, a process which allowed a species to expand its distribution (Grinnell, 1922). Brown (1984, 1995; Brown *et al.*, 1995) offers a niche based explanation to explain this pattern, arguing that the high abundance in the range core represents the location where the various biotic and abiotic conditions, i.e. niches axes, are most conductive to a species' survival and long term persistence. Thus, as the similarity of sites decays with distance, the abundance distribution across the range can be seen to reflect the response of populations to local environmental conditions, and the degree to which these conditions meet the niche requirements of particular species (Brown, 1995). This theory also postulates that as limiting biotic and abiotic factors are generally not continuously or uniformly distributed across the range, the general trend of a decrease in abundance towards the range edge will not follow a perfectly linear relationship. The theory, therefore, provides an explanation for the random peaks and gaps in abundance often observed across the range (Brown, 1984; Gaston, 1990). This niche based hypothesis can be visualised by thinking in terms of optimal resource surfaces: the favourable resource conditions for a species are located in the range centre and these decrease towards the range margins (Gaston, 2003).

An alternative hypothesis is a dispersal based metapopulation model. This follows a more general trend in ecology in which niche based models have 'fallen out of favour,' as a consequence of increased theoretical discussion in the literature surrounding neutral and dispersal based models. This metapopulation model is based on source-sink dynamics and

hypothesises that individuals are distributed as a result of dispersal from a source area, rather than as a response to local environmental conditions (Hanski, 1999). In such a system source areas are generally preferable habitat for a particular species, in which recruitment rates exceed the death rate, and are more likely to be located in the range core. Individuals emigrate from the source area to less favourable sink patches towards the range margins, in which death rates exceed birth rates and population density is reduced; thus, immigration is responsible for maintaining population persistence (Pulliam, 1988; Brown, 1995; Hanski, 1999). Finally, Bell (2001) argues that neutral theory can also explain this variation in abundance across the range.

Confirmation of the core-abundance hypothesis has not been universal. Sagarin & Gaines conducted a meta-analysis of 145 separate studies and found only 39% supported the core-abundance hypothesis. However, the authors do list potential methodological shortfalls, such as lack of sampling sites located in the range periphery that may have influenced detection of the pattern in the original studies. Other authors have also failed to find strong evidence of the pattern (e.g. Rapoport, 1982; Blackburn *et al.*, 1999).

The core-abundance hypothesis provides two potential explanations for H1. First, it predicts that as resources and habitat quality, in reference to a given species, decline towards the range margin, extinction rates are increased as one approaches the range edge (Sagarin & Gaines, 2002; Gaston, 2003). Thus the critical island area is expected to be higher (i.e. a larger island area) due to niche-based reasons. Whilst the change in critical island area across a species range has not been explicitly examined before, a number of incidence function studies have observed similar patterns. For example, Royle & Kéry (2007; see also Karanth *et al.*, 2006) found for the cerulean warbler (*Setophaga cerulea*) in the USA that there was an increased probability of turnover and greater instability in site incidence (i.e. a reduced survival probability) towards the edge of the range; an observation Royle & Kéry partly

attributed to less suitable niche conditions in range edge sites. To take another example, Svensson (1992) analysed patterns of occupancy, abundance and niche breadth in three beetle species in Sweden. It was observed that two of the species conformed to the core-abundance hypothesis, in that abundance, occupancy and niche breadth all decreased as the range edge was approached (Svensson, 1992).

A second more trivial explanation for H1 can also be derived from the core-abundance hypothesis. If abundance does decline from the range core to the range edge: i) random placement models predict that just by chance a species will occur in fewer patches in a landscape, and these are more likely to be the larger patches (Coleman *et al.*, 1982; Gaston, 2003), and ii) fewer individuals of a given species are expected to be present in any one fragment at the range edge and thus the possibility of missing a species during sampling is increased (Gaston & Blackburn, 2000). Both of these possibilities mean that the CIV will be higher for range edge landscapes relative to the values observed in landscapes located in the range core. As all three of these possibilities (i.e. the niche based explanations, and the two random placement/sampling explanations) involve decreasing abundance from the range core to the range edge, determining the exact process driving varying CIVs across the range is problematic.

2.8 Macroecological patterns of species abundance

I have published three review papers on species abundance distributions (SADs) during my doctoral study, and due to constraints on word limits for the thesis I have decided to amalgamate two of these three reviews into one condensed section. I have also expanded parts of the published reviews in order to present a more comprehensive review of a variety of macroecological patterns of abundance relevant to the thesis.

The two condensed papers are:

Matthews, T.J. & Whittaker, R.J. (2014a) On the species abundance distribution in applied ecology and biodiversity management. *Journal of Applied Ecology,* doi: 10.1111/1365-2664.12380

Matthews, T.J. & Whittaker, R.J. (2014b) Species abundance distributions: a review of recent advances in approaches for fitting and comparing competing models. *Frontiers of Biogeography,* **6,** 67-82.

The third paper focuses on the SAD in the context of neutral theory, and is less relevant to the core aim of thesis. Nonetheless, it does incorporate interesting information and has thus been included as Appendix 1. The reference for this paper is:

Matthews, T.J. & Whittaker, R.J. (2014c) Neutral theory and the species abundance distribution: recent developments and prospects for unifying niche and neutral perspectives. *Ecology and Evolution*, **4,** 2263-2277.

2.8.1 The species abundance distribution

> We are optimistic that the species abundance distribution will, in the long run, prove to be one of the most important and fundamental tools in community ecology and management.
>
> (McGill *et al.,* 2007, p. 1010)

2.8.1.1 Species abundance distributions: the basics

A species abundance distribution (SAD) describes the abundance of all species recorded within an ecological community, assemblage or sample. The study of SADs has a long

history in biogeography and community ecology (e.g. Preston, 1948, Tokeshi, 1993, McGill *et al.,* 2007), and the observation that ecological systems contain few very abundant species and relatively more rare species (the so-called hollow curve) is often labelled as one of the few universal laws in this subject space (McGill *et al.*, 2007, Ulrich *et al.,* 2010). SADs also provide the theoretical foundation for exploration of other biogeographical and macroecological patterns, such as the species–area relationship (e.g. Preston, 1948, Hubbell 2001, Whittaker & Fernández-Palacios, 2007).

A plethora of SAD models has been described in the literature; indeed, a review by McGill *et al.,* (2007) lists 27 different models. A summary of the models referenced in this review is given in Table 2.4. These models can largely be classified based on the mechanisms underlying the model, and traditionally models have been dichotomised as either statistical or biological (i.e. non-niche-oriented and niche-oriented). The distinction between statistical and biological models, as Magurran (2004, p.28) states, has become 'blurred,' and many distributions originally classified as statistical, such as the logseries and lognormal, have since been interpreted biologically. For example, niche-based theories concerning resource allocation have predicted logseries and truncated lognormal distributions (Table 2.4; Sugihara, 1980, Connolly *et al.,* 2005), while demographic models based on ecological drift (such as Hubbell's interpretation of neutral theory, see Appendix 1) predict logseries SADs at the metacommunity scale. As an aside, several authors (e.g. Tokeshi, 1993, Magurran, 2004) further classify SAD models into deterministic and stochastic models and stress the importance of knowing where a model falls in this dichotomy, in terms of fitting. In practice, the former include many of the statistical models, such as the lognormal and logseries, and make the assumption that individuals are distributed between species in a deterministic manner; that is, frequency values cannot vary. Stochastic models, which include the entire set of niche-oriented biological models with the exception of the geometric series (Table 2.4),

incorporate a degree of randomness and assume that abundance will vary between communities structured through the same set of rules—thus allowing these models to better incorporate random variation such as may arise from sampling (Tokeshi, 1993, Magurran, 2004). The characterisation of a particular model as deterministic or stochastic has important implications when testing the fit of the model with real ecological data.

Table 2.4 Descriptions of the various species abundance distribution (SAD) models referenced within the review. For each model, a description, selection of relevant references, and the model family according to McGill *et al.,* (2007) are given.

SAD model	Family	Description	Relevant references
Gambin	Purely statistical	Gambin is a stochastic model which combines the gamma distribution with a binomial sampling method. Gambin has a single free parameter (α), which characterises the distribution shape.	Ugland *et al.* (2007)
Lognormal	Purely statistical	The lognormal represents a situation in which the logarithms of the different species' abundances follow a Gaussian distribution and as such it characterises a community with relatively few very abundant or very rare species; whereby the observed frequencies increase to a modal frequency and then decrease.	Preston (1948), May (1975)
Truncated lognormal	Purely statistical	As for the lognormal, but the distribution is truncated at the left hand side in accordance with Preston's (1948) concept of the veil line: the idea that rarer species are often missed because of under-sampling and thus that the left hand side of the lognormal distribution is not revealed. This truncation methodology has since been criticised as inexact and other authors have advocated techniques such as Poisson sampling and hypergeometric sampling, as more accurate alternatives (e.g. Dewdney 1998).	Preston (1948), Dewdney (1998)
Logseries	Purely statistical	The logseries (logarithmic series) results from the Poisson sampling of a gamma distribution after a certain relevant limit is taken, and conditional presence is considered. That is, it gives the conditional probability of attaining a certain abundance level given that the species is present. The number of species at different levels of abundance predicted by the logseries is given by: $S_n = (\alpha * x^n)/n$, where α and x are constants, and S_n is the number of species with n individuals. This distribution is characterised by a skewed J-curve with a modal value of 1 in arithmetic space. There	Fisher *et al.* (1943)

		is no fixed variance assumption and an increase in sample size leads to an increased variation in log abundance.	
Poisson lognormal (PLN)	Purely statistical	An issue with the continuous lognormal is that it allows fractional abundances and does not have an associated sampling theory. One workaround in this regard is to incorporate Poisson sampling variability and fit the Poisson lognormal distribution: the Poisson sampling of individuals from a standard lognormal distribution.	Bulmer (1974)
Niche partioning models	Niche partioning	Models in which it is assumed that the total niche space of a community, often termed 'the stick', is sequentially divided in some manner between the species that comprise this community. Species' abundances are then related to the proportion of the stick acquired by each species. The development of such models has a long history, starting with Motomura's (1932, see Numata *et al.*, 1950, for a biological interpretation) geometric series, and includes MacArthur's (1957, 1960) three niche apportionment models. Of these, the broken-stick model is of particular interest in that it is one of the only SAD models to be rejected by its developer. More recently, Tokeshi (e.g. 1993, 1999) proposed several niche partitioning models.	Motomura (1932, see Numata *et al.* 1950), MacArthur (1957, 1960), Sugihara (1980), Tokeshi (1993, 1999)
Zero-sum multinomial distribution (ZSM)	Population dynamics	The SAD predicted for the local community in Hubbell's (2001) spatially implicit neutral model. The ZSM has three parameters: the number of individuals in the local community (J), the probability that a dead individual is replaced by a randomly selected immigrant from the metacommunity (m), and the 'fundamental biodiversity number' (θ). The zero-sum assumption is simply that when an individual dies it is immediately replaced by another individual, i.e. resources are fully saturated at all times.	Hubbell (2001), Matthews & Whittaker (2014c)

The easiest way to interpret SADs is to plot them. This produces graphical output which can be used to determine the form of the SAD, and as a basic means of evaluation of the goodness of fit. Two main methods are employed to plot SADs. First, a histogram (frequency distribution) of the species' abundances can be constructed (Fig. 2.6a). The number of individuals (horizontal axis) can be untransformed or logarithmically transformed before being 'binned'. The binning of data into abundance octaves is widespread in the SAD literature (e.g. Preston, 1948, Volkov *et al.*, 2003; see Fig. 2.6a). Data can be binned in various ways; however, all result in the loss of information. Based on MacArthur (1957),

Whittaker (e.g. 1965) developed an alternative plotting method (variously termed 'rank–abundance Plots,' 'Whittaker plots,' and 'dominance–diversity graphs'; see Fig.2.5b) to circumvent the issue of losing information when binning. Rank–abundance plots are constructed through plotting abundance (untransformed or log transformed) against rank order, where rank one corresponds to the species with the highest abundance, rank two corresponds to the species with the second highest abundance, and so on. As a method to determine the best SAD model for any given data set, rank–abundance plots are not perhaps as intuitive as frequency distributions, but they are beneficial in highlighting differences in evenness between datasets, as well as not losing information in binning. Other SAD plotting methods exist, such as the empirical cumulative distribution function (e.g. McGill, 2011), but are rarely used in practice.

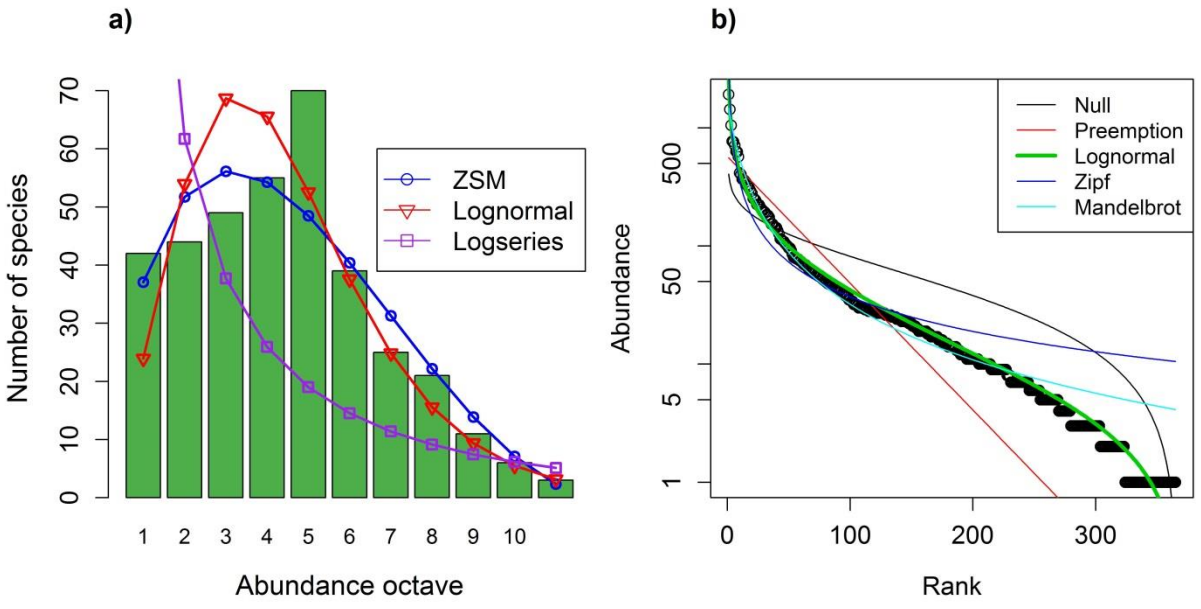


Figure 2.6 Exemplar SAD model fits illustrating the two major plotting methods: a) histograms and b) rank–abundance plots. In (a) the lognormal distribution (red line and triangles), logseries distribution (purple line and squares), and zero-sum multinomial distribution of Hubbell's (2001) spatially implicit neutral model (blue line and circles) have been fitted to binned simulated data (green bars, 365 species and 22945 individuals) derived

through the random sampling of individuals from a simulated lognormal metacommunity, and fitted using maximum-likelihood methods. The predicted value of the logseries model for the first octave has been omitted for clarity. In (b) abundance is ordered in decreasing magnitude and plotted against the corresponding rank in this order. The broken-stick model ('Null', black line), geometric series ('Pre-emption', red line), log normal (green line), Zipf (dark blue line) and Mandelbrot (light blue line) distributions have been fitted to simulated data (open circles).

2.8.1.2 Multimodal SADs

While the possibility has long been recognised, it has become increasingly apparent that many SADs may in fact be multimodal; that is, characterised by multiple distinct modes (e.g. Pielou, 1969; Dornelas & Connolly, 2008; Vergnon *et al.*, 2012). A number of explanations have been put forward to explain multimodal SADs, including emergent neutrality theory (Vergnon *et al.*, 2012; but see Barabás *et al.*, 2013) and the possibility that they simply represent statistical artefacts (Gaston, 1994; McGill *et al.*, 2007). A more detailed history of multimodal SAD research is presented in Table 2.5. However, perhaps the most widely accepted explanation is the aforementioned amalgamation hypothesis (e.g. Alonso *et al.*, 2008). In this regard, analysis of multimodal SADs often reveals interesting information relating to community structure because the different modes in the distribution have often been found to represent clusters of different types of species (Ugland & Gray, 1982; Borges *et al.*, 2008). That is, species are more similar, according to some trait, to others within clusters than between clusters. For instance, one assemblage division which has proved enlightening in SAD studies has been into core and satellite species subsets. Conceptually, core species represent the constituent members of any ecological community and are predicted to be structured according to traditional niche-based mechanisms (Ulrich & Zalewski, 2006). Satellite species are those species only occasionally found in any community (i.e. mostly immigrants from outside the local species pool) and are predicted to be governed by stochastic processes, largely random dispersal. Using a core–satellite division Magurran & Henderson (2003), who focused on an estuarine fish community in the UK over

a period of 21 sampling years, discovered that those species which are relatively abundant throughout the record and possess specialised estuarine habitat requirements (the 'core' species) are characterised by a lognormal distribution. In contrast, the species with low abundances, infrequent records in the dataset and different habitat requirements (the 'satellite' species) follow simple Poisson processes and are characterised by a logseries distribution (Magurran & Henderson 2003). The two different sets of species leave different signals within the SAD, and when combined result in a distribution with more rare species than predicted by the standard lognormal model (for further examples see Gray *et al.,* 2005; Ulrich & Zalewski, 2006; Unterseher *et al.,* 2011). It is important to remember, however, that any division of an assemblage into core and satellite species is a simplification of reality. Empirical assemblages represent a continuum of species types and workers must be careful not to introduce artefacts into analyses when splitting species into binary categories.

Table 2.5 The history of multimodal species abundance distribution research.

Study	Main finding
Pielou (1969); Šizling *et al.* (2009)	One of the first studies to recognise the possibility of multimodal SADs, stating that increasing a sample to include a broader range of taxa (e.g. expanding from a focus on warblers to all birds), might result in the combination of several logseries distributions, each with its own set of parameters. More recently, Šizling *et al.* (2009) have discussed how incorporating multiple areas and a variety of taxa within a sample can result in a SAD that is a proportional sum of different partial SADs.
Ugland & Gray (1982)	A problem with using the lognormal distribution to model SADs is that it assumes equal density probabilities across the species (symmetry), but in reality there is a general pattern of asymmetry as communities are comprised of three distinct abundance groups, roughly translated as rare, intermediately abundant and common. Within each group the SAD is symmetric, which leads to a mixture of three lognormal distributions when focusing on the whole assemblage.
Gray *et al.* (2005)	Found strong evidence of bimodal lognormal distributions for a mixture of marine and terrestrial datasets.
Dornelas & Connolly	Fitted a combination of one to four mode Poisson lognormal (PLN) distributions, in addition to the logseries, for a large sample of coral communities, Australia,

(2008)	and compared the various distributions using maximum likelihood methods. At the scale of the entire sample the three-mode PLN had the greatest support, but as the four-mode model was within an AIC of 2 of the three-mode model this can also be concluded to have strong support.
Borges *et al.* (2008)	The SADs of arthropod communities on Terceira Island, Azores, could be deconstructed into three distinct abundance groups (abundant, intermediate and rare).
Dornelas *et al.* (2009)	Found that environmental heterogeneity underpinned multimodal SADs of weed seed bank communities.
Vergnon *et al.* (2012)	Tested a number of datasets representing various taxa and found strong evidence of multimodality. This indicates that multimodality may be more common than previously thought: it has simply been overlooked.

2.8.1.3 Biological SAD models

Biological models are largely niche-oriented models in which it is assumed that the total niche space of the community, often termed 'the stick', is sequentially divided in some manner between the species that comprise this community (Marquet *et al.*, 2003, Magurran, 2004). The analogy of 'the stick' was introduced by MacArthur (e.g. 1957, 1960), who proposed a number of different biological SAD models, including the broken-stick model (Table 2.4). MacArthur argued that if the focus was ecologically similar species competing in a community, a structured pattern of relative abundance results. According to MacArthur's model, niche partitioning can be estimated as a random partition of a common limiting resource (breaking the stick). The relative abundances of the species are proportional to the total fraction of the resources each species appropriates. However, this is just one particular model (i.e. set of rules for linking abundance with niche space), and many others have been proposed (Table 2.4). Within all biological models, species that then immigrate into the community must apportion whatever parts of the stick remain and break such sticks according to the rules of the particular model (see Fig. 2.7a for an illustration of this process). In this way, biological models are mechanistic, providing ecological rationale to the way in

which abundance is proportioned across species. This modus operandi assumes a one-to-one relationship between species' niches and species' abundances, a correspondence which is hard to test in practice, especially considering that a species has both a fundamental and (more restricted) realised niche (Hutchinson, 1957, Sugihara, 1980, Tokeshi, 1993). Nonetheless, this assumption is backed up by certain studies (e.g. Marquet *et al.*, 2003). Niche models are also largely zero-sum models in the sense that the number of individuals within the local community is generally fixed; an increase in the number of individuals of one species must equate to a decrease in the number of individuals in one or more other species. Several authors have advised removing the very rare species from a sample when calculating niche-based biological models (e.g. Tokeshi, 1993, Magurran, 2004, but see Fesl, 2002). This suggestion is valid in theory but requires careful consideration of what constitutes rarity at the level to be omitted.

Tests of niche models have produced mixed results (Tokeshi, 1990, Fesl, 2002, Mouillot & Wilson, 2002). Fesl (2002) compared a number of niche-oriented models, including six of Tokeshi's (Fig. 2.7b), using data sourced from larval Chironomidae assemblages in the River Danube, Austria. Interestingly, Fesl's study found that none of the models provided a good fit to the data when all species were included. However, when the whole assemblage was deconstructed into functional subgroups it was found that the 'random fraction model' (RAM: one of Tokeshi's models in which species apportion available niche space in a random manner; Fig. 2.7b) generally provided the best fit amongst functional groups. Strong evidence also existed for the 'random assortment model' (another of Tokeshi's models in which abundances vary independently of one another; see Tokeshi, 1990, 1993). Based on the latter finding it was deduced that, in this system, species' abundances are largely independent of each other (Fesl, 2002). The RAM was also found to

provide the best fit to a fish parasite community, again indicating that species' abundances vary independently (Mouillot *et al.*, 2003).

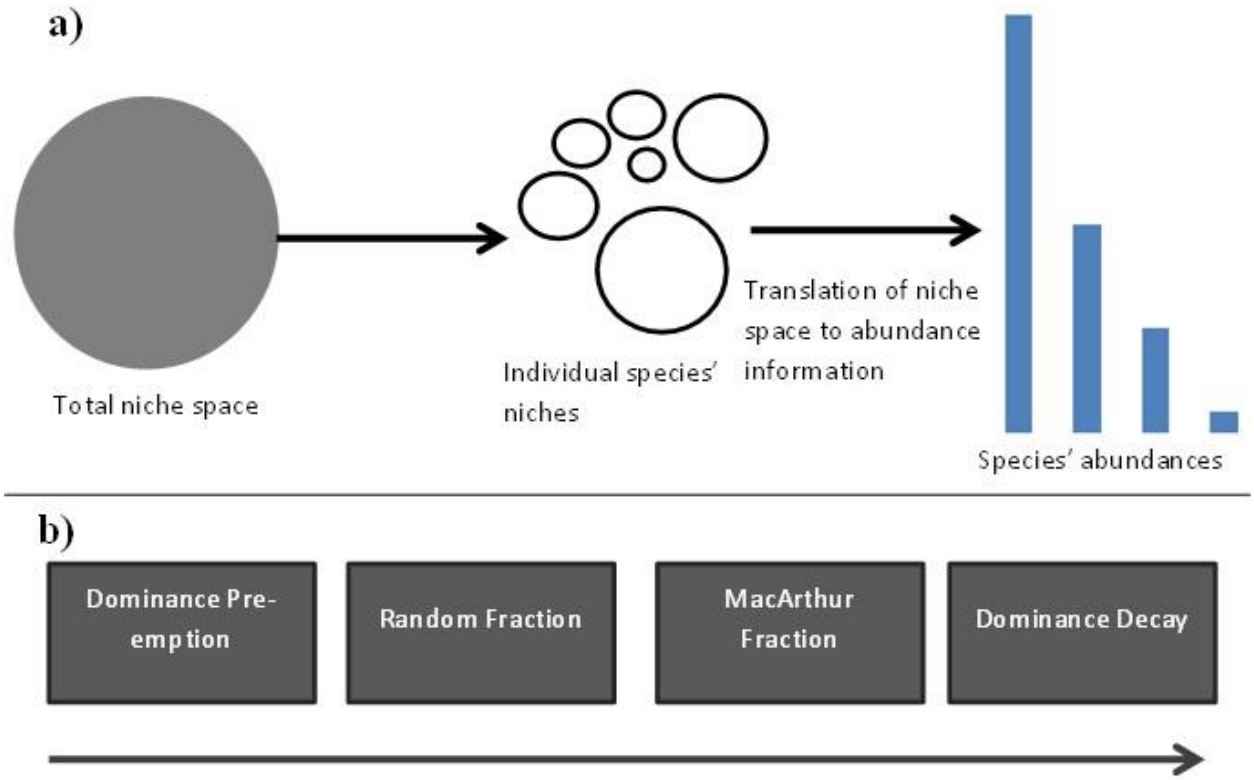


Figure 2.7 Niche-oriented species abundance distribution models, where (a) illustrates the conceptual basis of niche-oriented SAD models, and (b) sets out the spectrum of a selection of Tokeshi's sequential-breakage niche models. These are models in which it is assumed that the total niche space of the community, often termed 'the stick', is sequentially divided between the species that 'invade' the community according to a set of rules. Going from left to right the arrow represents the increase in evenness/uniformity of species' abundances in the community. The Dominance Pre-emption Model (Tokeshi, 1990, 1993) predicts a fixed dominance hierarchy and can be seen as conceptually similar to the geometric series (Table 2.4). The first species, the superior competitor, pre-empts more than 50% of the total niche. The second species, the next best competitor, then apportions more than 50% of the remaining niche space, and so on. The inverse of this division is the Dominance-decay Model (Tokeshi, 1990), whereby it is always the largest niche in the community that is divided; thus, the probability of a niche being apportioned is related to the niche size. The MacArthur Fraction Model (Tokeshi, 1990, 1993) specifies a sequential process of niche subdivision in which total niche space is randomly divided in two, with one of these resulting niches further chosen (again, this is based on the idea that the niche space of the most abundant species is most likely to be invaded) and divided randomly. The Random fraction model is similar, except that it assumes that niches of varying size have the same probability of invasion. See Tokeshi (1990, 1993, 1999) for a full description of these and additional niche-oriented SAD models. Part (a) is adapted from Tokeshi (1993).

2.8.1.4 On the SAD in applied ecology

Although not as widespread in the management and conservation literature as other macroecological patterns (e.g. the SAR), SADs are a potentially useful tool for conservation scientists and managers as they can be used to 'define' rarity in a particular community. Conservation practices generally focus on rare species within a delineated protected area boundary (Gaston, 1994) and a SAD can provide evidence of the level of rarity of particular species of interest relative to other species (McGill, 2011), and thus extinction risk and associated conservation action. In addition, changes in the empirical SAD can act as an early warning for the effects of disturbance on biodiversity, as the shape of the SAD can change markedly before any local extinctions occur (Hågvar, 1994; Mouillot *et al.*, 2013). Disturbance to ecological communities arising from drivers such as habitat loss, invasive species, and pollution is a pervasive feature of modern times (Sala *et al.*, 2000), and there is thus an exigent need for the development of tools that allow for relatively quick assessments of ecosystem health and/or the success of management prescriptions aimed at ameliorating the effects of disturbance (Mouillot *et al.*, 2013).

I focus here primarily on the use of the SAD in disturbance ecology as it is the area of research most relevant to focus of the thesis (i.e. the loss and fragmentation of natural habitat), but see Matthews & Whittaker (2014a) for a more comprehensive review of the different applications of SADs in applied ecology (e.g. fisheries management, conservation planning).

Traditionally, the lognormal distribution has been proposed as a means of accurately modelling undisturbed communities. This assertion is based on the Central Limit Theorem, in which a large number of biotic and abiotic factors acting multiplicatively generate a lognormal distribution of abundances (May, 1975). In contrast, it has been argued that more disturbed systems should be characterised by a logseries SAD (e.g. Gray & Mirza, 1979; Hill

et al., 1995). This transition from lognormal to logseries shaped curves is illustrated in Fig. 2.8. Here, I have simulated an undisturbed community to follow a lognormal distribution (Fig. 2.8a). I then modelled a natural perturbation to the system by randomly sampling 5% of the individuals (i.e. removing 95% of the individuals) (Fig. 2.8b). Ecologically, this simulated perturbation represents a situation in which populations of rare species are most likely to go locally extinct. This simulation is also similar to the approached adopted by Green & Plotkin (2007), in which individuals were sampled in a spatially explicit manner from regional SADs to examine the scaling relationships between sample and regional-scale SADs. Plotting the SAD of the undisturbed and disturbed community (i.e. before and after) reveals the shift from a lognormal to a logseries distribution (Fig. 2.8). Comparing the fit of the Poisson lognormal and logseries distributions provides a more rigorous means of determining deviation from the lognormal and using Akaike's information criterion (AIC) to compare the two models reveals that in Fig. 2.8a the lognormal provides a superior fit, whilst in Fig. 2.8b the logseries is the better-fitting model (AIC values are given in Fig. 2.8).

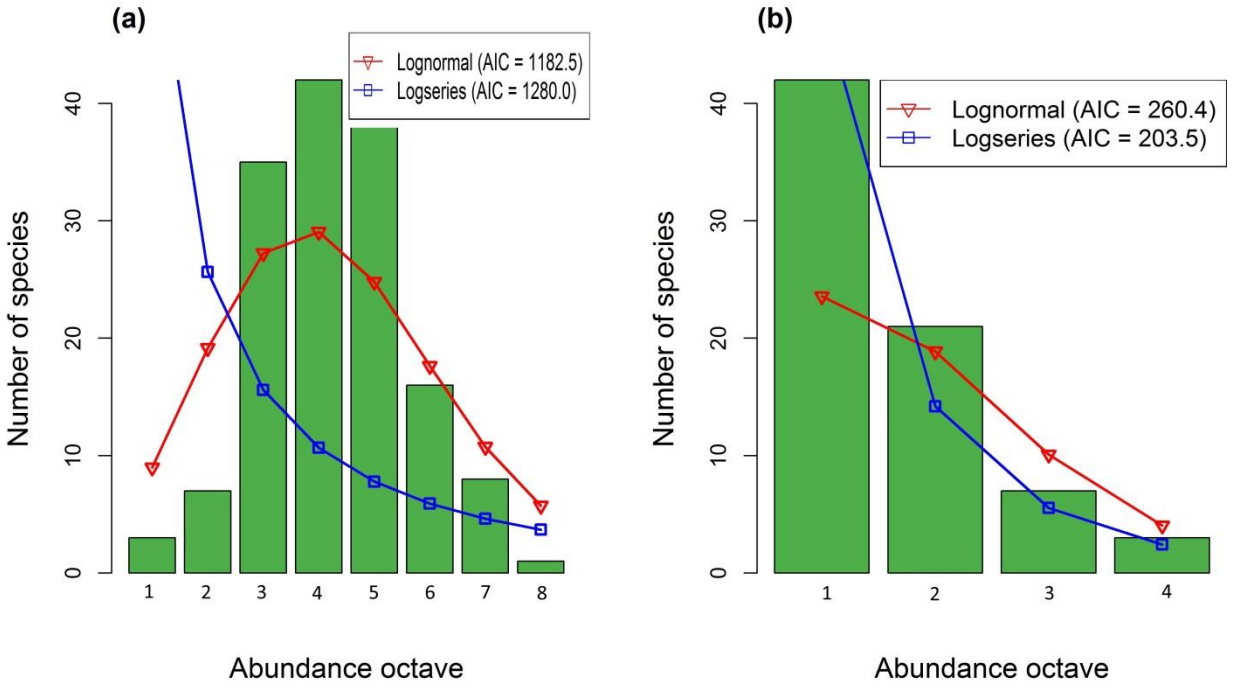


Figure 2.8 Using the species abundance distribution to determine the effect of disturbance on biotic community structure. Panel (a) represents an undisturbed community, and the data were simulated by sampling from a log-normal distribution (N = 3000; S = 150). We then applied a perturbation (i.e. a disturbance event) to this community, by randomly sampling 5% of the individuals (i.e. randomly removing 95% of individuals). This resulted in a more logseries-like distribution of abundances (panel b; N = 150; S = 73). In each instance, we fit the Poisson log-normal distribution (PLN; zero-truncated version) and logseries distribution to the data and calculated the Akaike's information criterion (AIC) for the fit of each distribution. To calculate AIC, the distributions were fitted to the raw abundance data (i.e. not binned). The data were binned for graphical purposes only. The predicted values on the plots (a,b) have been generated from fitting the models to the binned data; the predicted value of the first octave from the logseries distribution has been omitted to increase clarity. In (a), the PLN clearly provides a superior fit, while in (b), the logseries has a lower AIC.

It is important to note that the use of the lognormal in such circumstances is based on the assumption that it provides a good fit to the undisturbed data, but other distributions, including the logseries, have often been found to fit data from undisturbed systems better (e.g. Syrek *et al.*, 2006), while disturbed communities have been shown to follow a lognormal distribution with similar parameters to undisturbed communities (Nummelin, 1998). Furthermore, under acute stress, communities have been observed to go from logseries back to lognormal, but with very different parameters (Hågvar, 1994). Due to these issues,

other SAD models have been proposed as more accurate descriptors of undisturbed and disturbed communities, e.g. the Zipf–Mandelbrot model (Mouillot & Lepretre, 2000), geometric series (Kim *et al.,* 2013), broken stick and negative binomial distribution models (Syrek *et al.,* 2006), and power law and niche partioning models (Tang *et al.,* 2010). Thus, fitting a suite of models to both undisturbed and disturbed communities can be informative (e.g. Syrek *et al.,* 2006). Additionally, instead of attempting to detect deviation from lognormal to logseries distributions, the aforementioned multimodal SAD models can be used to measure the impact of disturbance. For example, Dornelas *et al.,* (2009) showed that for agricultural weed communities, homogenous environments were characterised by a standard unimodal lognormal SAD. However, weed communities in increasingly heterogeneous environments (e.g. due to tillage and nitrogen fertilisation) exhibited increasingly multimodal SADs (Dornelas *et al.,* 2009). A further issue relates to the fact that the SAD of a poorly sampled assemblage (i.e. a sample which represents a small proportion of the number of individuals in the assemblage) may be mistaken for a disturbed community's SAD. This is because in the situation where individuals are randomly distributed across space, small samples from lognormal-like assemblage SADs can result in truncated lognormal curve shapes (Green & Plotkin, 2007). Thus, it is important to attempt to keep sampling intensity constant between the communities being studied.

Disturbance as an ecological concept is wide ranging and need not be confined to pollution. Habitat loss and fragmentation are also forms of ecological disturbance that may impact on community properties, including the SAD, which in turn may offer a rapid assessment tool (Hill *et al.,* 1995; but see Nummelin, 1998). The aforementioned deconstruction approach would be beneficial in this endeavour as the SAD should become more logseries-like with increasing fragmentation. However, and as with the SAR (above), an

influx of generalist and matrix species into recently fragmented habitat may modify or obscure changes in the shape of the SAD.

Histograms are not the only method for plotting the SAR and Fig. 2.9 provides a graphical summary of the various SAD methods available for examining the impact of disturbance and management actions on ecological communities. For example, Lambshead *et al.,* (1983) pioneered the K dominance plot (Fig. 2.9a). This method plots cumulative abundance percentages against species rank and has been argued to be a useful way of determining the effect of disturbance on the SAD, but has been little explored. In K dominance plots, if a curve lies completely below another curve (as community A does in Fig. 2.9a), then this community can be defined as being more diverse. A similar method derived in the context of marine benthic communities, uses *K* dominance plots to compare the distribution of individuals among species, with that of biomass (Fig. 2.9b; e.g. Warwick 1986). It is based on the theoretical consideration that the standard individual-based SAD should behave differently from the biomass-based SAD when affected by disturbance. Unpolluted systems are characterised by the biomass curve appearing above the abundance curve, as in Fig. 2.9b, and *vice versa*. Methods based on departures from a lognormal SAD shape (using histograms) and comparisons of model parameters (Fig. 2.9c, d) have been discussed above. The empirical cumulative distribution function (ECDF) can be plotted instead of the commonly used probability density function, and can be useful at highlighting differences in SADs between communities (Fig. 2.9e). For instance, three hypothetical SADs (here, we have formulated the data ourselves for effect) are plotted in ECDF form in Fig. 3e, each relating to a river invertebrate community under different pollution scenarios: high and mildly polluted, and unpolluted. In (e), the unpolluted curve lies below the two polluted curves at the left hand side of the plot, indicating that there are higher proportions of low abundance species in the two polluted com- munities. Finally, the slopes of SAD models

plotted using rank abundance diagrams (RAD) offer an alternative to histograms for comparing SADs between communities (see Fig. 3f).

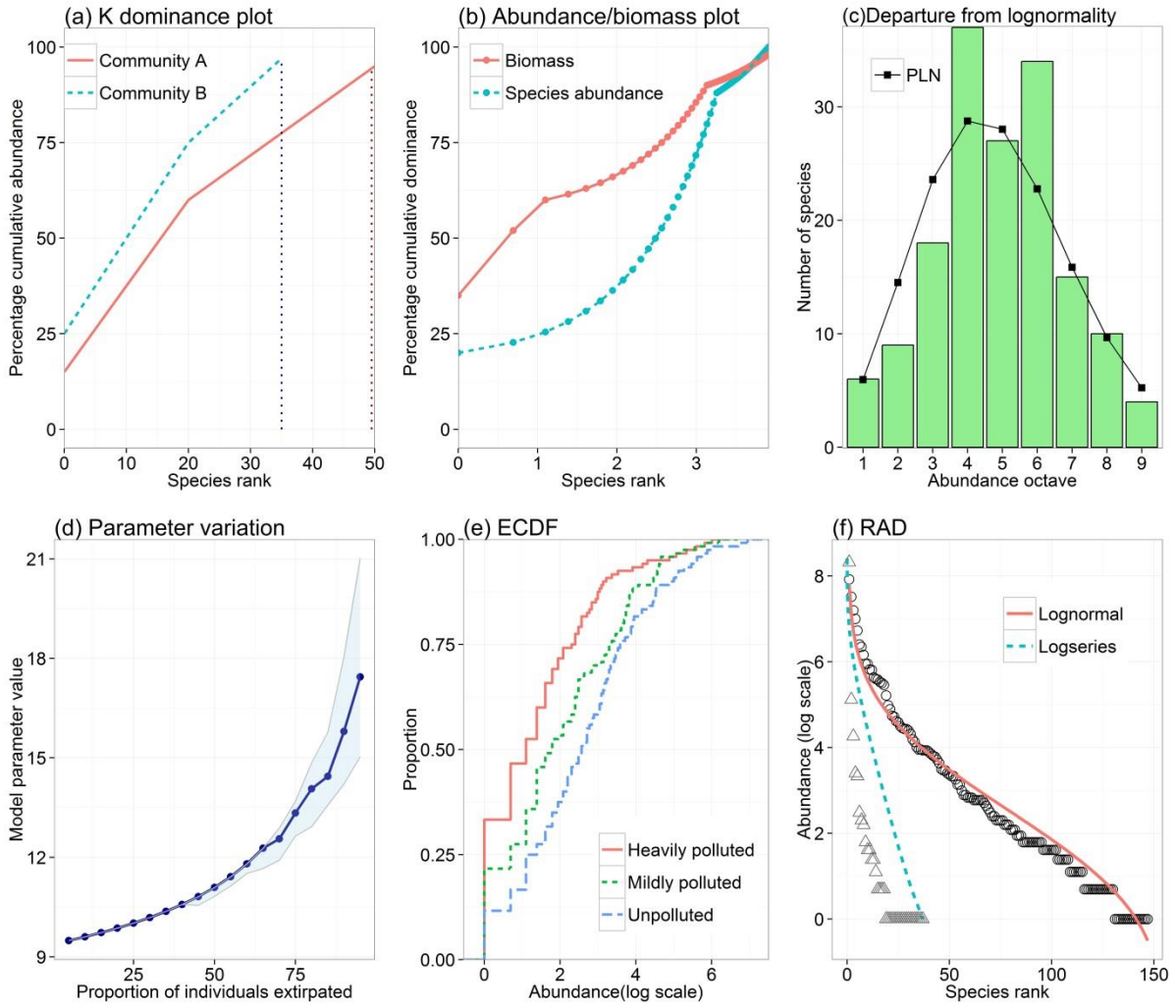


Figure 2.9 Species abundance distribution-based methods for determining the effect of disturbance in applied ecology using simulated (a–e) and empirical (f) data. (a) A K dominance plot (i.e. a plot of cumulative abundance percentages against species rank; Lambshead, Platt & Shaw 1983). If a curve lies completely below another curve [as community A does in (a)], then this community can be defined as being more diverse. The vertical dashed lines in (a) indicate the species richness of each community. (b) Abundance/biomass plots (i.e. a K dominance plot that compares the distribution of individuals amongst species with that of biomass; see Warwick 1986). Unpolluted systems are characterized by the biomass curve appearing above the abundance curve, as in (b), and vice versa. (c) The fit of the log-normal (PLN; here, the zero-truncated form) to an empirical SAD (here, binned) is often used to assess the impact of disturbance on species assemblages. (d) Variation in the logseries alpha along a hypothetical disturbance gradient in which an increasing proportion of the individuals from a community (simulated to follow a log-normal SAD; N = 2396, S = 52) are randomly lost through sampling from the community (left to right on the x-axis represents increasing hypothetical disturbance and loss of individuals). Sampling was repeated 100 times (dots = median alpha values; shading = 95% confidence

intervals). (e) Three hypothetical SADs plotted using the empirical cumulative distribution function (ECDF). The data were created for this figure in order to illustrate the method, but might for instance relate to river invertebrate communities subjected to different degrees of pollution: high and mildly polluted, and unpolluted. In (e), the unpolluted curve lies below the two polluted curves at the left hand side of the plot, indicating that there are a higher proportion of low abundance species in the polluted communities. (f) The SADs of the pooled polluted (triangles) and unpolluted (circles) Norwegian marine invertebrate samples (see Fig. 2. caption) are plotted in rank abundance form. The fit of the truncated log-normal model to the unpolluted samples, and the logseries model to the polluted data, are displayed. In the polluted sample, it is clear that there are more rare species predicted even by the logseries. The plots were constructed using the ggplot2 R package (Wickham 2009).

2.8.2 Correlation between abundance and range size

Another commonly reported abundance-distribution pattern is the positive correlation between both global and local abundance of a species (untransformed or log transformed), and various measures of that species' range size (Andrewartha & Birch, 1954; Brown & Maurer, 1989; Gaston, 1990; Gaston & Lawton, 1990; Blackburn *et al.,* 1997; Gaston & Blackburn, 2000; He *et al.,* 2002; Gaston *et al.,* 2006; see Hanski & Gyllenberg, 1993, 1997 for discussions in a metapopulation context; see Fig. 2.10 for an example using number of grid squares occupied as a metric of range size). Put another way, common species can be seen to have broad ranges while rare species can be seen to have comparatively restricted range sizes. For instance, in an analysis of 67 British bird species abundance was found to explain around two-thirds of the variation in range size (Gaston & Blackburn, 2003). The pattern, also known as the occupancy-abundance relationship (He *et al.,* 2002; Gaston & Blackburn, 2003), has been found for variety of ecosystems and multiple taxa including, *inter alia,* plants (Guo *et al.,* 2000), birds (Gaston & Blackburn, 2000), bees (Anasiewicz, 1971), mammals (Blackburn *et al.,* 1997), and marine species (Foggo *et al.,* 2003); however, it is only apparent when comparing ecologically similar species, such as guilds (Gaston & Lawton, 1990). It also appears to be constant between native and invasive species (Labra *et al.,* 2005). Using the US Breeding Bird Survey data Labra *et al.,* discovered that the relationship between range size and abundance did not significantly differ between natives

and exotics, despite the observation that exotic species obtained higher max abundances. The relationship between abundance and distribution also holds at multiple spatial scales, although global abundance generally has a stronger relationship than local abundance (Gaston, 1990; Bell, 2001), and the slope of the relationship itself is scale specific (Borregaard & Rahbek, 2010). This point of scale independence is of particular interest as it suggests that the correlation may indicate the constraints of different ecological relationships operating at different spatial scales. The shape of the relationship has generally received less attention than for other macroecological patterns, such as the SAR, but generally follows a power law when considering global abundance (Bell, 2001).

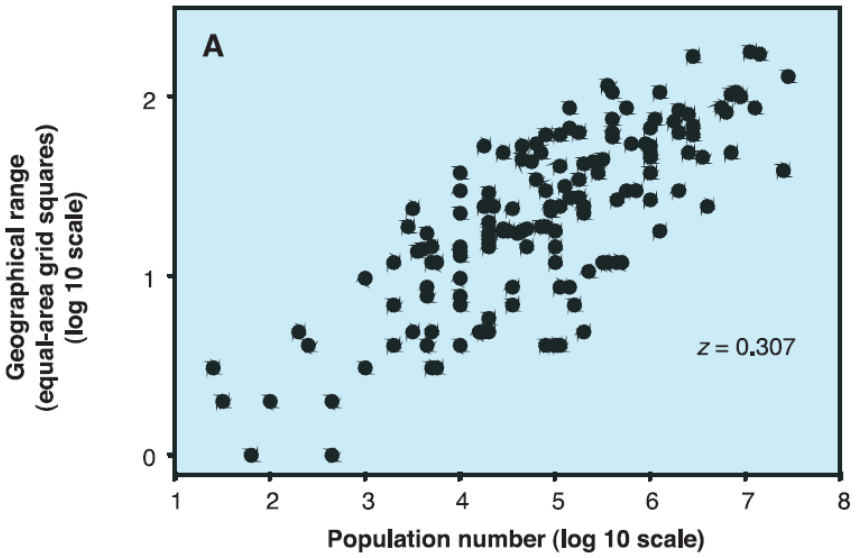


Figure 2.10 The relationship between global abundance and global range size (number of grid squares occupied) for Anseriformes. Taken from Bell (2001).

Before progressing to the possible mechanisms behind this abundance-distribution correlation, it necessary to highlight that although highly prevalent it is, as with most ecological 'laws' not a universal macroecological pattern. For example, a study by Rabinowitz *et al.*, (1986) categorised British plants into eight categories (seven implying rarity of some sort) based on abundance, habitat specificity, and range size. The researchers

found examples of species that fit into almost all categories; the only exception being the category incorporating a large range with broad habitat specificity and low abundance. In a general sense this lack of universality is due to the complexity of ecological systems, the multitude of different environmental variables that interact and influence species at multiple spatial and temporal scales, and the sheer volume of factors acting to determine distributions. Reviewing the studies that have researched this lack of universality reveals several main areas of focus. One such focus has been the influence of colonisation (Gaston & Blackburn, 2003); metapopulation models predicting species with high colonisation rates should have greater levels of occupancy for their abundance and *vice versa* (Hanski, 1994a; Hanski, 1999; Gaston & Blackburn, 2003). Gaston & Blackburn (2003) investigated whether divergence from the positive abundance-occupancy relationship could be explained by differences in dispersal, an element of colonisation. However, despite the published observation that dispersal ability and both occupancy and abundance are correlated (Gaston, 1994; Gaston & Blackburn, 2003), it was found that including metrics on natal and adult dispersal explained little of the extra variation in occupancy that is not already explained by abundance (Gaston & Blackburn, 2003).

When reviewing the literature on ecological patterns of this nature, and in particular the explanations offered as the mechanism behind the pattern, it can often be confusing trying to distinguish between model types. Firstly, there are statistical models and distributions proposed to *model* the relationship, of which many are not based in any ecological rationale. Then there are true biological hypotheses proposed to *explain* why a pattern occurs and to determine the mechanisms behind deviations from the common trend. Obviously this is by no means a true dichotomy as numerous statistical distributions are based on ecological theory and many theories predict a certain distribution. Nonetheless, bearing in mind this difference, numerous statistical models have been proposed to describe the occupancy-abundance

relationship including models derived from the Poisson distribution, the negative binomial distribution - dealing with the aggregation of species distributions in space, Nachman's model, Hanski & Gyllenberg's (1997) logistic model, and He *et al.*'s general model of which the other models are special cases (see He *et al.,* 2002, for a review of the above models and full references). At the same time a huge variety of hypotheses have been proposed to explain the pattern including sampling, structure and dynamic based theories, and readers are directed to a comprehensive review by Borregaard & Rahbek (2010) for a full list.

During the 1970s and 1980s the generally accepted explanation of the abundance-occupancy pattern theorised that certain species possessed particular evolutionary adaptations offering them competitive advantage, which led to the establishment of broad ranges and large population sizes (Brown, 1984; Bock, 1987). For instance, an oft cited hypothesis published during this period is the resource use hypothesis (sensu Borregaard & Rahbek, 2010) of Brown (1984, 1995). This comprises a niche based explanation for the pattern, outlining that small differences between ecologically similar species allow certain species to have large ranges and increased abundance across this range. Correspondingly, species with a narrow niche are only present in habitats that meet their specialist requirements. As with the niche-based explanation for decreasing abundance towards the range periphery discussed above, this theory is centred on the assumption that the abundance and distribution of species reflects the ability of individuals to satisfy their limiting biotic and abiotic needs. However, empirical evidence of the resource use hypothesis is mixed (e.g. see Gregory & Gaston, 2000, against; and Heino, 2005, for).

Other authors have argued that the pattern can be better explained through a focus on metapopulation dynamics, namely, dispersal (particularly the rescue effect), colonisation and extinction dynamics (e.g. Hanski *et al.,* 1993; Gonzalez *et al.,* 1998). Hanski's (1982) core-satellite hypothesis is one such example, postulating a bimodal frequency distribution of

species in terms of the number of sites where they are present. In simple terms the ‘core species’ are those with high abundances that are present in numerous sites, whereas the ‘satellite species’ are only located in a modicum of sites, and are generally characterised by low abundance (Hanski, 1982). As the highly abundant ‘core species’ are widely distributed they possess a greater chance of producing emigrants to sites where they are already present, and thus increasing the long term persistence of populations through the rescue effect (Brown & Kodric-Brown, 1977), and to sites where they are not present, thus enabling the species to increase its distribution through colonisation (Hanski, 1999). In addition, as ‘satellite species’ have a relatively restricted distribution and low abundance they produce fewer emigrants, leading to reduced colonisation and population persistence. As a result of this circularity the ‘core species’ remain abundant and widely distributed, and the ‘satellite species’ remain rare. The metapopulation dynamics hypothesis is realistically only relevant at landscape and regional scales. As Gaston & Blackburn (2000, p.169) note “abundance-range size relationships observed at much larger scales, such as that shown by birds across the whole of Britain, are unlikely to be generated by a mechanism based on interpatch dispersal.”

In a similar vein to sampling arguments often levied at SAR studies, this scaling of range size with abundance could potentially be the result of a sampling artefact, namely that if a species is relatively abundant it is more likely to be sampled across its whole range than a relatively rare species (Gaston, 1990; Wright, 1991). As such, Wright (1991) contests that simply observing the pattern is ecologically ‘meaningless’ and cannot be used to validate any biological hypotheses. Furthermore, the phylogenetic non-independence of species is an additional statistical issue which has been put forward as a potential mechanism. However, while true in theory, a wealth of empirical evidence (see Gaston & Lawton, 1990; Gaston & Blackburn, 2000) suggests that the sampling artefact and phylogenetic non-independence issue are unlikely to be the sole causes behind the abundance-range size pattern.

Other potential hypotheses include:

- Range position – species abundance and patch occupancy have been found in certain instances to decrease from the range core to the periphery (Brown, 1995). If this relationship is a true ecological pattern it may help to explain occupancy-abundance relationship. This is because the species whose range margins are located in a particular region will be expected to have lower abundance and lower occupancy. However, numerous studies have found no evidence of the core-abundance pattern (e.g. Rapoport, 1982; Blackburn *et al.*, 1999), and Gaston *et al.*, (1997a) found that it cannot fully explain occupancy-abundance relationships.
- Niche position/resource availability – species which use resources that are numerous and widespread can be expected to be numerous and widespread themselves (Gaston & Blackburn, 2000). Gregory & Gaston (2000) have found support for this hypothesis in the occupancy-abundance relationship of British breeding birds.
- Vital rates – if r equals a species' birth rate minus its density-independent death rate, the species with larger r will be able to obtain higher abundances and occupy a greater number of sites (Holt *et al.*, 1997; Gaston & Blackburn, 2000).
- A final group of hypotheses deal with spatial aggregation and postulate that distribution-abundance relationships are a result of the clumped nature of the distributions of individuals (for further discussion, see Borregaard & Rahbek, 2010).

Finally, the literature review up to point has not discussed any aspect of the issue of causality in the context of the abundance-distribution relationship, that is, does abundance drive distribution or *vice versa*, is an external factor responsible for controlling the relationship, or is there mutual causality between these two variables through time (Borregaard & Rahbek, 2010). Clarifying the direction of causality is not simply an

interesting academic exercise; rather there are a number of potential conservation implications. For instance, it is possible to conceive a situation where the abundance of a species is reduced through overexploitation. It would thus be of some significance for managers responsible for the 'maintenance' of the species to have a sense of the likely effect, if any, of this overexploitation on the species' distribution. Notwithstanding the above implications that flow from causality, and irrespective that certain authors appear to assume that causality is 'bidirectional' (Bock, 1987; Borregaard & Rahbek, 2010, p.2), a review by Borregaard and Rahbek (2010) suggests that no empirical evidence exists in favour of any particular causality relationship.

A criticism often directed at ecologists is that while they are proficient at theorising multiple hypotheses for any given pattern, rarely do they come to a consensus on what is driving a particular observation (Smith, 1997). Discussion in this section up to to this point has done little to disprove this view. Rather, it has identified the widely researched abundance-distribution pattern accompanied by a plethora of mostly conflicting explanatory theories. The thrust of this section is primarily to review the relevant literature, and the revealed lack of intellectual consensus is perhaps inevitable, given the complex nature of the biological world and the likely fact that there is no one hypothesis responsible for most ecological patterns. What is clear is that while no evidence provides unequivocal support for any one individual hypothesis (Gaston *et al.*, 1997b; Borregaard & Rahbek, 2010) the causality of the relationship between distribution and abundance is a complex interaction of a variety of factors operating at different scales and affecting individual species differently, and it is likely that the different hypotheses are not mutually exclusive, rather it is a combination of processes that act to produce the patterns discussed herein.

CHAPTER 3: METHODS

3.1 Preamble

The empirical component of my thesis is split into two parts. Firstly, I conducted a number of meta-analyses based on a large number of datasets sourced from the literature (Chapters 4-11). Secondly, I undertook an analysis based on data I obtained from my own field work (Chapter 12). The divide between the two components is blurred as my field work datasets were used in most of the meta-analyses. However, for ease of explanation I have outlined the methods for each component separately in the following sections.

The exact methods, including the statistical and analytical methods, computer programs used, sensitivity analyses and method justification, are given in each of the empirical chapters. In this chapter I have simply summarised how the data (secondary data and field work data) were gathered and organised; but not analysed.

3.2 Data collection from the literature

In order to answer many of the thesis objectives it was necessary to source a large number of habitat island datasets from the literature. The first step involved looking for datasets in which a taxon had been sampled across a set of habitat islands of varying area; all types of habitat island were considered. This initial search incorporated two types of dataset: (1) datasets in which the only data available were the number of species of a given taxon on each island, and (2) sources which provided the full species list for each island. Both types of dataset were stored for use in particular analyses (e.g. type 1 and 2 datasets were used in Chapters 4 & 5, whereas just type 2 datasets were used in Chapters 6 & 7).

Three abstracting services were searched for relevant habitat loss and fragmentation studies and datasets during May 2011 – May 2014: JSTOR (1913–2003); ISI Web of Knowledge (1980–2014); and BIOSIS Biological Abstracts (1980–2003). The search

keywords were variants on ‘habitat islands and species richness’ and ‘forest fragments and species richness’. Additionally, cross-referenced papers derived from the reference lists of sourced papers were also included. Primarily, I only included a dataset if the data were fully published in the source paper, that is, I did not generally contact the source paper authors for data. In a few cases the data were provided by the authors of the source papers; but ultimately these datasets were not used in the analyses, due to issues with co-authorship etc.

All datasets sourced by the above steps were then subjected to a set of stricter criteria in order to determine whether they were suitable for inclusion (criteria 1 & 2 relate to the initial dataset search but are included here for completeness):

1. Habitat islands were defined as discrete patches of habitat surrounded by contrasting, ‘hostile’ matrix habitat.
2. The areas and species richness of each habitat island were listed.
3. The number of habitat islands within each dataset was at least four, as this was necessary so that all the various analyses could be computed.
4. Data from any particular study did not overlap with those from any other study accepted for analysis (data for different taxa within the same study system were allowed).
5. The sampling methods were deemed to be sufficient. This is inherently a subjective criterion but care was taken to ensure that appropriate sampling methods had been used. I read the methods sections of each source paper, and only accepted datasets which had been published (i.e. the methods had been peer-reviewed). Whilst this is not a foolproof option, I am confident that datasets based on very poor sampling were excluded. Nonetheless, the potential issues regarding data quality in meta-analyses are discussed in Chapter 13.

6. The sampling of a given taxon in a set of habitat islands was the primary aim of the study. This criterion was introduced in order to exclude studies in which the sampling of habitat islands was a secondary objective. For example, I came across studies which were interested in carbon sequestration in forest fragments, and which also reported the number of tree species. However, it was not clear that the authors had been rigorous in determining the number of tree species in a fragment as this information was not essential to their study. Thus, these studies were excluded. Criteria 5 and 6 clearly overlap to some degree, but I found it useful to consider them separately when evaluating a dataset.

For each dataset I recorded a large number of dataset characteristics: the geographic location, the latitude and longitude of the study extent, the taxon studied, the habitat island type, the range of species richness and island sizes, and any other variables that were given in the source papers (e.g. isolation, disturbance indices). All island areas were converted into hectares. The habitat island type was recorded in two ways: (1) as specific as possible (e.g. deciduous temperate forest fragments), and (2) in order to create a categorical variable which could be used in regression models, as one of a smaller range of coarser categories. These categories were as follows: forest islands, grassland fragments, mountain-top islands (sensu Brown, 1971), urban fragments, islands within lakes, and other (e.g. gravel pits). Forest island datasets were defined as those in which the main vegetation type was forest, and the surrounding matrix was non-urban. Urban fragment datasets could encompass any main vegetation type (typically it was forest), where the surrounding matrix was predominantly comprised of urban land uses.

Where possible, for each dataset I also recorded the predominant land-use type of the matrix surrounding the habitat islands using information in the source papers. Where possible I checked this information using Google Earth Pro (Google Earth, 2013), although this check was undertaken tentatively as many datasets were compiled decades ago and it is likely that the land-use has changed in the interim. These land-use categories were then converted into a 5-point scale of matrix permeability in which 1 represented the most permeable matrices, and 5 represented the least permeable matrices. This classification scale was partly based on the results of Chapter 5: a study which examined how the slope of the island species–area relationship varied across multiple habitat island types. Obviously, this is a coarse approach, but obtaining fine scale matrix data for such a large number of studies is unfeasible, particularly as such information is rarely, if ever, presented in the source papers. The matrix type was classified as:

1 – Secondary/re-growth forest matrix (if the island type was forest), or a scrub matrix (if the island type was grassland)

2 – Secondary/re-growth forest matrix (if the island type was grassland), or a scrub matrix (if the island type was forest)

3 – Agricultural matrix (e.g. crops, pasture)

4 – All mountaintop islands were attributed to this category as the matrix between them is known to be relatively impermeable (Brown, 1971; Lomolino *et al.*, 1989; see Chapter 5)

5 – Urban land-uses and water matrices (i.e. flooded forest fragment archipelagos)

This information was stored in a database I created to manage the data. In total, I sourced 229 datasets, 104 of which included full species lists. These 229 datasets comprised

138 vertebrate, 51 invertebrate, and 40 plant datasets; and 154 forest, 14 mountaintop, 17 grassland, 22 urban and 22 'other' habitat island datasets. The full dataset information and references are presented in the various result chapters and associated appendices. For the analyses in Chapter 5, I also sourced the 601 true island (e.g. oceanic islands) datasets used in Triantis *et al.* (2012). These were obtained from one of the authors of Triantis *et al.* (Francois Guilhaumon) and re-analysed. Similar dataset selection criteria were used in gathering these datasets and readers are directed to Triantis *et al.* (2012) for the full methodology and dataset summaries.

The following sections summarise the fieldwork methods I used to obtain bird-habitat island data for four fragmented landscapes in Europe. As stated above, these data were primarily used to construct incidence functions in Chapter 12, but I also included them as datasets in a number of the meta-analyses.

3.3 Field site and patch selection

The aim of my Chapter 12 analyses was to determine whether a species' incidence function varies in a predictable manner across the species' range. Thus, to test this hypothesis it was necessary to select landscapes in the range core and range edge of certain species. First, a set of target woodland species were selected with roughly overlapping ranges in Europe (see section 3.5.2 below for details of the target species). Second, I looked for suitable sampling regions in the core and edge of these species' ranges. Ideally, I would have been able to sample a large number of landscapes along a gradient from the range core to the range edge (see discussion in Chapter 13). However, due to logistical constraints, I opted for two core and two edge landscapes. The actual landscapes to be sampled were selected by searching the literature for studies which had undertaken some sort of habitat island sampling within the region (e.g. Aparicio, 2008; Jamoneau *et al.*, 2011). This meant that I was able to contact the

authors of these papers for useful information (discussed below). Potential landscapes were then examined for suitability using Google Earth Pro (Google Earth, 2013). Landscapes were only selected on the basis that the predominant land use type was agricultural, as matrix type has been found to influence species incidence functions (Watson *et al.*, 2005). Figure 3.1 & 3.2 illustrate the location of these study landscapes. The exact location of the four landscapes was (the latitude-longitude co-ordinates are the approximate co-ordinates of the centre of the landscape extent):

- North France (n=40 fragments; 50° 8'52.11"N, 1°40'37.09"E)
- South Spain (n=42 fragments; 36°11'31.16"N, 5°30'13.61"W)
- Central Norway (n=41; 59°28'0.60"N, 11°21'12.97"E)
- South UK (n=37; 51° 7'56.51"N, 0°34'57.84"W)

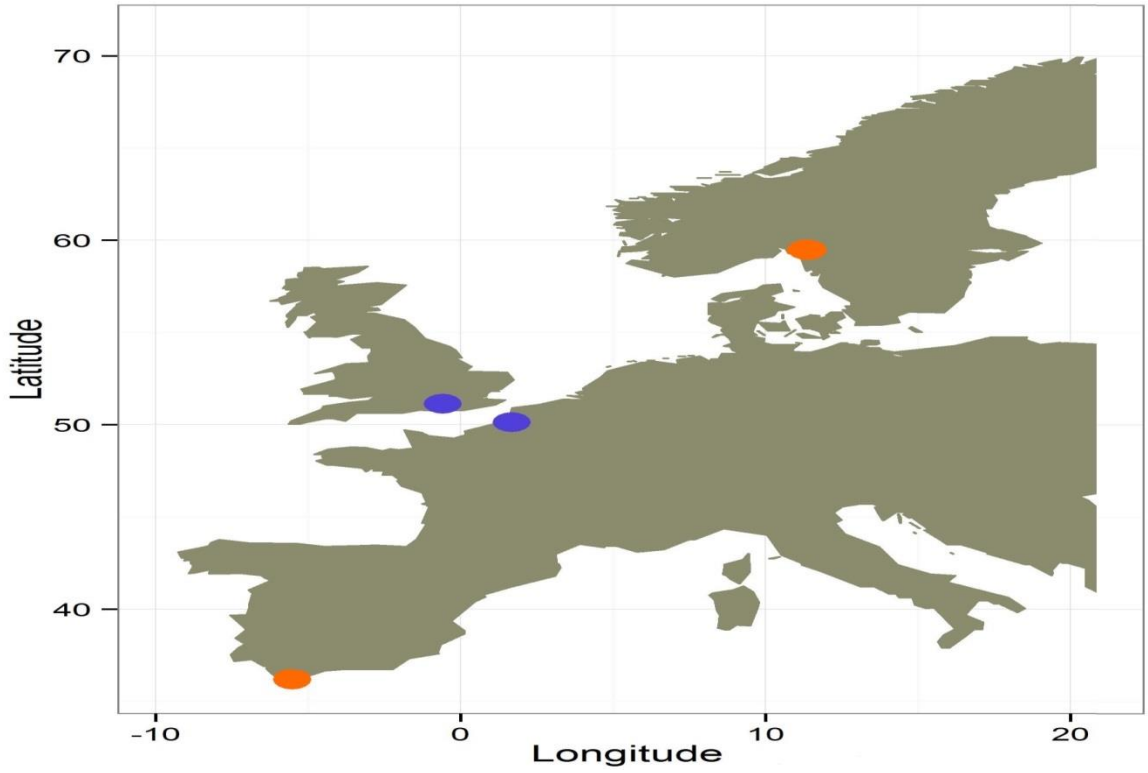


Figure 3.1 A map of Europe with my study landscapes highlighted. Blue dots represent the range core sites, and orange dots represent the range edge sites. See Fig. 3.2 for more precise locations.

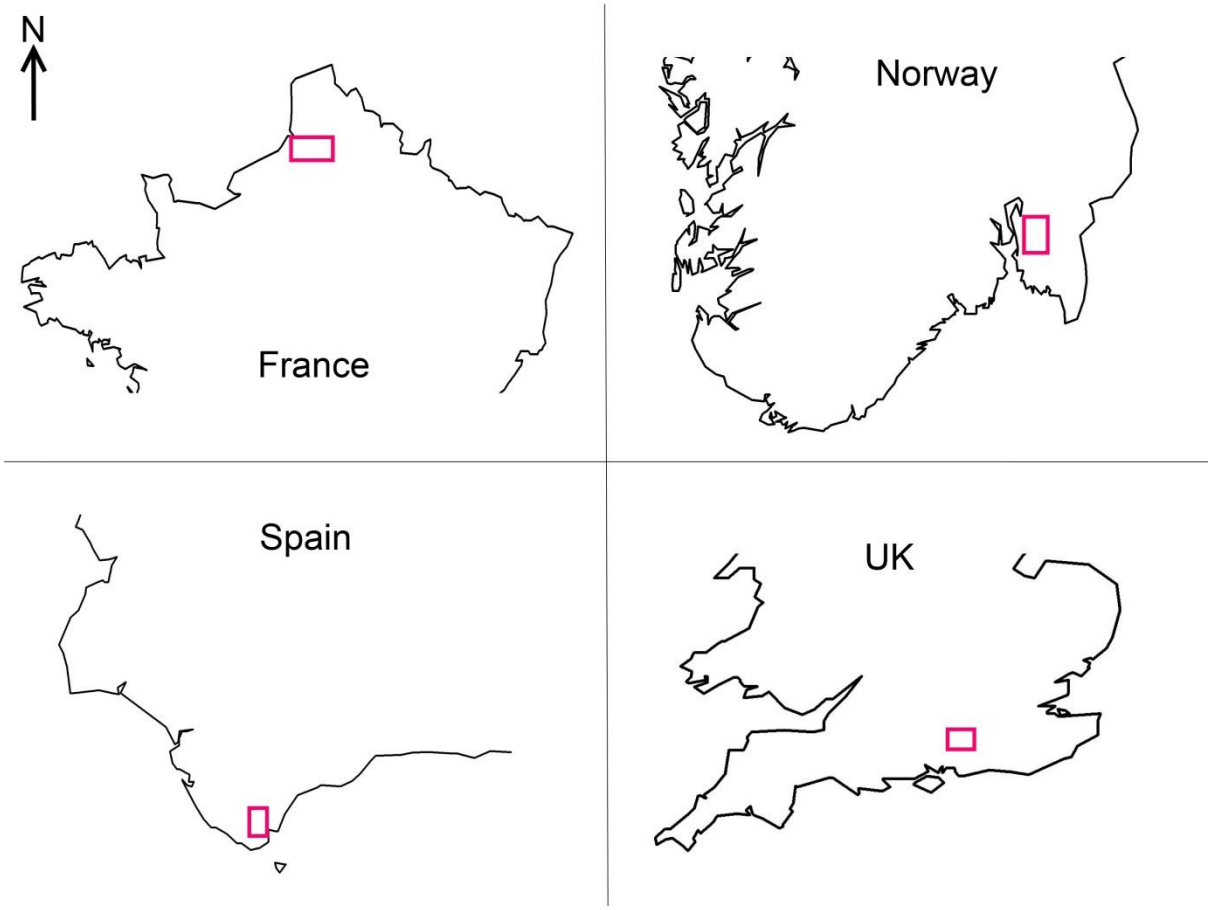


Figure 3.2 The location of my four study landscapes (red square) within the relevant countries. The individual plots are drawn to scale, but for reasons of presentation the four together are not to scale.

The French landscape was located in North France (Fig. 3.2) near the town of Abbeville. The area is rural and the predominant land use type is agricultural, dominated by open fields growing cereals and sugar beet. A number of small forest fragments persist in the landscape (e.g. Fig. 3.3), owned largely by farmers and private landowners. The Forêt de Crécy, a 4300ha public forest is the nearest "continuous" forest. The dominant vegetation is broadleaf temperate forest, dominated by English oak (*Quercus robur*) and European beech (*Fagus sylvatica*). The mean temperature and average annual rainfall are 10 °C and 700 mm, respectively (Jamoneau *et al.*, 2011).



Figure 3.3 An example study site in north France: a forest fragment isolated by an agricultural matrix.

The Spanish landscape was located in south Spain (Fig. 3.2); north and west of the city of Algeciras in the Andalusia region. The landscape has long been managed for intensive agriculture (Aparicio, 2008), and a number of crops are grown, including cereals, olives and legumes. Eucalyptus plantations are also common. The remaining natural vegetation is largely Mediterranean forest dominated by cork oak (*Quercus suber*) (Fig. 3.4). However, historical stone pine (*Pinus pinea*) plantations are extensive and are now considered semi-natural (Corona *et al.,* 1988); most fragments contained some stone pine. Mean annual temperature is around 18°C, and annual precipitation ranges from 460 to 1,020 mm (Aparicio, 2008).

The Norwegian landscape was located next to the town of Rakkestad in south east Norway (Fig. 3.2). The surrounding municipality is one of Norway's largest agricultural areas and the landscape contains a mixture of grain fields and pasture land (dairy farming). Forestry is also a common land use. Fragments of native forest remain and represent a mixture of Boreal and Nemoral vegetation, dominated by Norway spruce (*Picea abies*), silver birch (Betula pendula), and aspen (*Populus tremula*), with a large number of other species

present, including English oak (*Quercus robur*), ash (*Fraxinus excelsior*), and Scots pine (*Pinus sylvestris*). Mean annual temperature ranges from 3°C to 11°C, and annual precipitation is around 750mm a year (World Weather Online, 2014).

Finally, the UK landscape was located in south England in the county of Surrey. Again, the landscape has long been dominated by intensive arable agriculture. The native vegetation is primary temperate broad leaf forest, dominated by English oak (*Quercus robur*), although in certain areas the soil is more acidic and coniferous species, such as Scots pine (*Pinus sylvestris*), are more prevalent. Mean annual temperature ranges from 6°C to 15°C, and annual precipitation is around 650mm a year (Met Office, 2014).



Figure 3.4 A fragmented cork oak landscape in Andalusia, south Spain, with me in the foreground. Photo courtesy of Christie Palmer.

Within each landscape, individual fragments were chosen using a stratified random method. Due to time constraints, I was only able to sample approximately 40 fragments per landscape, and as my empirical analyses were focused on the importance of island area (e.g.

SARs, incidence functions) it was important that the fragments covered a wide range of areas. Thus, I formulated a number of area classes and selected fragments at random with an aim to distribute the areas across the classes. The classes were as follows: 0.01-4.99ha, 5.00-9.99ha, 10.00-14.99ha, 15.00-19.99ha, 20-29.99ha, 30.00-49.99ha, and 50ha+ (area measurements are described in section 3.4, below). As the largest first derivatives in SAR curves and incidence function models occur at relatively small to medium island areas (i.e. curves tend to flatten out at relatively large areas; Arrhenius, 1921; Rosenzweig, 1995), the classes were designed in such a way to allow for more fragments in the smaller area classes. I attempted to select only fragments which were isolated from surrounding vegetation. However, this was not always possible as few forest fragments in these types of landscapes are completely isolated; many are connected by hedgerows, treelines or small forest-type corridors.

Sampling took place during April–July over two years (2012 & 2013). Each landscape was only sampled once; although each fragment within a landscape was sampled up to three times. Every patch was sampled at least twice, but due to time constraints the third round of sampling was not completed for every patch. Thus, when compiling the data after both field work seasons, species incidence records were only based on the first two rounds of sampling; although the third was used in the field to assess whether sampling was relatively complete. The landscapes in France and Spain were sampled in 2012, in that order, and the landscapes in Norway and the UK were sampled in 2013, again in that order. This approach was necessary as I conducted sampling on my own. To reduce bias regarding sampling certain landscapes earlier in the breeding season I sampled a core landscape first and an edge landscape second in 2012, and an edge landscape first and a core landscape second in 2013. Evidently, there are potential issues with only sampling a landscape in a single breeding season and these are discussed more thoroughly in Chapter 13.

With the exception of the UK, for each landscape I contacted a local ecological researcher (i.e. an academic from the country in which the landscape was located) for information on various factors, such as obtaining permission from landowners and potential accommodation etc. To set aside the maximum amount time for sampling in each landscape I arranged to stay in the landscape itself (e.g. Fig. 3.5).



Figure 3.5 My sampling accommodation in a mixed Boreal forest, Norway.

3.4 Patch characteristics

Patch area was measured by tracing round the fragments using ArcGIS (ESRI, 2011) and Google Earth Pro (Google Earth, 2013). As with any study of this nature, there is likely a degree of measurement error involved with measuring fragment area, particularly due to the aforementioned issue of connected tree lines etc. Isolation was measured in two ways using ArcGIS and Google Earth Pro (Google Earth, 2013). First, the Euclidean distance from the edge of the patch to the edge of the nearest patch (i.e. nearest neighbour distance) was recorded. Second, the edge to edge distance to the nearest patch larger than 10ha was also measured (cf. Watson *et al.,* 2005). Whilst it would have been preferable to include isolation

metrics based on the amount of habitat in a concentric circle buffer of a given radius around a patch (e.g. Bender *et al.*, 2003), this was not possible for all of my landscapes due to data availability. As isolation was not a primary variable of interest in my empirical analyses, I do not see this as a significant issue. Patch shape was measured using the shape index outlined in Patton (1975): SI = (Perimeter in feet)/(2*((Area in sqft * π)^0.5)). SI measures the deviation from circularity (Patton, 1975; Ewers & Didham, 2007).

For the two landscapes sampled during my first field work season (France and Spain) I also calculated the proportion of different land uses within a specified buffer width of each patch using ArcGIS (ESRI, 2011). For each landscape I imported a base-map of satellite imagery and traced round features using polygons to delineate the major land use types: arable land, woodland (including the study fragments), roads, rivers, lakes, plantations, scrub, and built up land uses. These polygons were then integrated into a raster using the mosaic tool, and the individual study fragment points were overlaid over this raster. A buffer was then created around each fragment point and the proportion of the buffer represented by each of the above land use classes was calculated using the raster buffer summary tool in Hawth's Analysis Tools (Beyer, 2004). Two buffer sizes were used (see Watson *et al.*, 2005): 1km and 2.5km. Unfortunately, I was unable to undertake this land use summary approach for my remaining two landscapes due to (a) a lack of suitable high quality satellite data, and (b) time constraints. Thus, I was only able to use these land use data in exploratory analyses, and they were not used in the main empirical analyses reported below.

3.5 Bird sampling methodology

3.5.1 Sampling method

Within each landscape the fragments were sampled for birds using ten-minute point counts. Whilst it would have been technically acceptable to use unbounded point counts as I was not

interested in measuring abundance, I used the preferable bounded point count method (see Fig. 3.6; Bibby *et al.,* 2000; Gibbons & Gregory, 2006). Originally, I had planned to record the distance from the centre of the point count to every individual seen or heard. However, in practice this was too difficult in the field and thus only very rough estimates were taken. Two boundaries were used: a 0-25m (from the centre of the point count) and 25-50m. Boundaries were estimated using a Leica Disto D510 Laser Distance Meter and specific landmarks (e.g. trees). In the smallest fragments (e.g. those under 1ha) the 50m and/or 25m point count boundaries extended outside the fragment. In these cases, all species seen or heard in the fragment were recorded, but species outside the fragment were not included. In certain fragments, the 50m boundary was difficult to estimate for all point counts, despite my best efforts. However, the 50m radius boundary was deemed necessary as the fragments were not very diverse, overall density was relatively low, and using just the 25m boundary resulted in much fewer individuals being sampled. This in turn would have required more point counts per fragment, for which I did not have the time to conduct. As I was only interested in species richness and incidence (i.e. not abundance), this issue is not that problematic (Gibbons & Gregory, 2006). The potential issue regarding variation in species detectability across the different landscapes is discussed in Chapter 12.

Counts were restricted to ten minutes as pilot studies (described below) indicated this to be a suitable time, the woods were not particularly diverse, I was familiar with the European avifauna, and it avoided the issue of double counting (Gibbons & Gregory, 2006). The number of point counts per fragment was proportional to the area of the fragment (Schoereder *et al.,* 2004), and the exact number of point counts was determined using a combination of pilot studies and recommendations from the literature: 1 point count for fragments < 5ha, 2 counts in fragments 5-9.99ha, 3 counts in fragments 10-14.99ha, 4 counts in fragments 15-19.99ha, and so on, although for the very largest fragments (i.e. those over

40ha) there were proportionally fewer point counts due to time constraints (see Hinsley *et al.*, 1995, 1996). Point counts were located systematically, such that there was a minimum of 200m between counts (Bibby *et al.*, 2000), and to ensure that both the patch core and edge were sampled (birds outside the fragment were not counted). Each fragment was sampled two-three times, with the exact time and day of the week of sampling randomised in each instance, and species accumulation curves were constructed to ensure sampling was relatively complete (Gotelli & Colwell, 2010). All birds seen or heard within the point count boundary were recorded. Individuals that were clearly transient (e.g. flying over the fragment) were noted as such and removed from all subsequent analyses. Counts took place from dawn to 10.30am, and were not conducted in bad weather (i.e. rain or high winds).

Prior to commencing sampling in each landscape a short pilot study was conducted. The duration of the pilot study varied between landscapes but was approximately five days. The pilot studies had a number of aims (Sutherland, 2006). First, it allowed me to determine whether the sampling methodology (e.g. length of point counts) was sufficient. Second, it allowed me to identify any potential issues, e.g. issues of land access. Third, it gave me a short amount of time to become accustomed with the local avifauna and to learn regional variations in songs and calls.



Figure 3.6 The view from a point count in a forest fragment in France.

3.5.2 Target species

As stated above, all species seen or heard within the point count boundaries were recorded. However, my Chapter 12 analyses were focused on a specific subset of target species. Specifically, these analyses involved fitting various incidence function models (see Chapter 2) to six target woodland species (Table 3.1). The target species were selected such that my two range core landscapes (i.e. France and south UK) were located in the these species' range core, and the range periphery landscapes (i.e. south Spain and Norway) were located in the edges of these species' ranges. For the purposes of this study, I classed the range as the breeding range of the species, which was taken by combining the distribution of the species in which it was present all year round with its summer only breeding range; the winter only range was excluded (Cramp, 1977; Gaston, 1994, 2003). In total there were six species which matched this criterion (Table 3.1).

Determining a real range edge is problematic, and as has been noted elsewhere, with a few exceptions (e.g. single island endemics) strict range borders do not exist (MacArthur, 1972; Blackburn *et al.*, 1999; Gaston, 2003). One issue is that a range edge which represents the edge of a landmass may not represent a "real" range boundary in the sense that the species is unable to survive/breed further out; it is simply that there is no available land. This is a particularly pertinent issue for my target species as the UK represents one such ambiguous range edge. For example, the range of the Eurasian blue tit (*Parus caeruleus*) is shown in Fig. 3.7. Here, it is clear that my Norwegian and south Spanish landscapes do represent landscapes near the edge of the species' range as the range could theoretically expand further (i.e. there is available land), but only does so to a small extent. In contrast, the distribution of the species extends right up to the tip of UK and would likely extend further north if land was available (Fig. 3.7). Thus, to be conservative I only considered a range edge to be "real" if there was a possibility for the species to be distributed beyond it. For example, the distribution of the Eurasian nuthatch (*Sitta europaea*) only extends up to north England (i.e. not Scotland) and this can be considered a "real" range edge (see Cramp & Perrins, 1993, p.300 for the distribution of the Eurasian nuthatch). For this reason, the Eurasian nuthatch was not considered one of my target species (see Table 3.1). For species whose distribution extended to the tip of the UK, I considered my south UK landscape to be located in the range core. Other studies have made similar suggestions; for example, Blackburn *et al.* (1999, p.496) considered "that the range centres of many birds in Britain arguably may be close to the range centre in Europe."

Range maps of the target species (Table 3.1) were obtained from the handbook of the birds of Europe, the Middle East and North Africa: the birds of the Western Palearctic (Cramp, 1985, 1988, 1992; Cramp & Perrins, 1993; Cramp *et al.*, 1994). The maps were digitised using ArcGIS (ESRI, 2011). There are many ways of drawing a range map

(Blackburn *et al.*, 1999; Gaston, 2003). Here, I traced round the edge of the range given in the source material (e.g. Fig. 3.7) to create a polygon in ArcGIS. To simplify matters I excluded island distributions that were far outside the continental range (e.g. the Canaries). I also truncated the range at the far eastern European/Asian border (roughly 50 degrees longitude) as this was the furthest point given in the birds of the Western Palearctic. This polygon tracing method meant I did not have neat concentric circles, but rather range polygons of complex form.

Following Channell & Lomolino (2001), I divided the range into two sections: the core and the edge. Deciding what actually constitutes the range core and edge is a subjective endeavour. To determine what constituted the edge of the range and what constituted the core range for my target species, I experimented with two criteria. First, I simply considered the area of the range within 1000km of a real boundary (above) as constituting the range edge. Secondly, I used a more strict definition based loosely on the approach of Channell & Lomolino (2000). As it is difficult to work with non-circular ranges, one work around is to convert the non-circular polygon into a circle of equivalent area. Then, using basic algebra, the radius of this circle can be calculated using the formula: $R1 = \sqrt{(A1/\pi)}$, where A1 = the area of this circle, and R1 the radius. I then reduced the area of the circle by a set amount, X (e.g. X = 25%). Using the above formula I calculated the radius of this smaller circle (R2), and then subtracted R2 from R1 to get a distance, D. D was then taken as the distance from a real range boundary that constituted the range edge. For example, if the area of the original range (A1) was 10,000km^2 (relatively small for a European bird species) then the radius (R1) would be 56.4km. Setting X to 25% gives a new radius (R2) of 46.4, and a D value of 10km. Thus, any site within 10km of a real range boundary would be considered an edge site. I experimented with two values of X, 10% and 25%. I ultimately chose 25% as the criterion value as 10% gave very small values of D (i.e. generally under 100km). Even using X= 25%,

this circle radius calculation always provided D values slightly less than 1000km. Thus, my Norwegian and Spanish study sites met both of these criteria for all of my target species (see Table 3.1), i.e. the 1000km criterion and the circle radius criterion.

Table 3.1 The target species for my field work sampling. *a photo of this species is provided as Fig. 3.8.

Common name	Latin name	Code
Eurasian blue tit	*Parus caeruleus*	1
Great spotted woodpecker*	*Dendrocopos major*	1
Common blackbird	*Turdus merula*	1
Eurasian wren	*Troglodytes troglodytes*	1
European robin	*Erithacus rubecula*	1
Wood pigeon	*Columba palumbus*	1

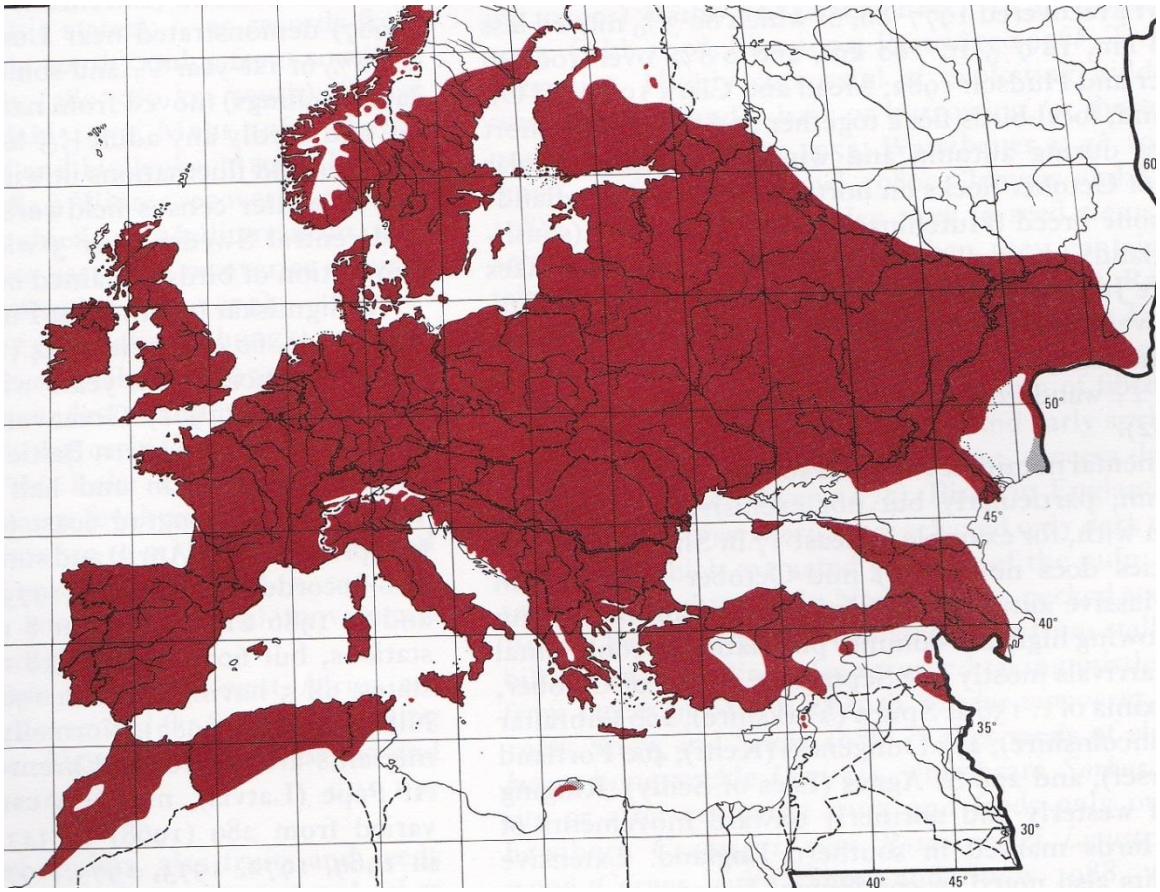


Figure 3.7 The range of the Eurasian blue tit (*Parus caeruleus*). Taken from Cramp & Perrins (1993).



Figure 3.8 The great spotted woodpecker (*Dendrocopos major*) – one of my target species. Photo courtesy of Steve Young.

3.6 Additional site information

Prior to commencing sampling I spoke with a number of ornithological experts and consulted the literature in order to select other patch variables which are known to be important in explaining bird distributions in fragmented landscapes (e.g. Hinsley *et al.*, 1994, 1995, 1996; Ford & Barrett, 1995; Bellamy *et al.*, 1996). These variables are summarised in Table 3.2. Vertical vegetation density was measured at each point count by looking directly upwards through a hollow cylinder (a kitchen roll: 45mm diameter) and estimating the percentage of the hole covered by vegetation. Horizontal density was measured using the board half obscured method (MacArthur & MacArthur, 1961). A 40cm by 40cm board constructed from plywood with black and white painted squares (10cm by 10cm) was constructed. At each point count the board was positioned around breast height where possible and I walked away

from the board in a straight line until 50% of the squares were obscured. The distance (i.e. the half sighting distance) was then measured using either a tape measure or a Leica Disto D510 Laser Distance Meter. For the vertical and horizontal density measurements, an average of the individual point count measurements was taken to compute a patch estimate for each variable. Within each point count boundary five canopy trees were randomly chosen and the diameter at breast height (DBH) of each noted. An average DBH was computed for the patch as a measure of patch maturity.

Table 3.2 The environmental variables recorded at each sampling site.

Variable	Method
Vertical vegetation density	A hollow cylinder was used to look directly upwards at the centre of a point count and the percentage cover of vegetation was recorded.
Horizontal vegetation density	The board half obscured method of MacArthur & MacArthur (1961; see main text).
Patch maturity	The average DBH of five randomly chosen trees within each point count boundary.
Habitat diversity	The number of habitat types recorded in a patch (see main text).
Habitat complexity	Determined using the rapid appraisal method developed by Newsome & Catling (1979).
Management information	Any relevant patch management information (e.g. an area of felled trees) was recorded.

Habitat diversity was measured by recording the number of habitats in a patch. This included habitat types observed in the point count boundaries, habitats observed on walks between counts, and during pilot studies. In addition, in each patch I traversed random transects in order to identify all major habitat types. Habitat types were adapted from the work of Hinsley *et al.* (1994, 1995, 1996) and were: deciduous woodland, coniferous woodland, mixed woodland, dense shrub layer, thin shrub layer, mature scrub (>2m), young scrub (< 2m), open glades, water features (pond, stream, river, ditch, drain), road or track

within wood, hedge as integral part of wood, and scrub. Finally, habitat complexity was measured using the rapid appraisal method of Newsome & Catling (1979) and Catling & Burt (1995). I followed Watson *et al.*'s (2009) implementation of the method. At each point count I constructed a $400m^2$ quadrat. Within each quadrat, five habitat complexity variables were graded following the classification in Newsome & Catling (1979) and Watson *et al.* (2009) (see Table 3.3). These five variables represented a range of habitat complexity measurements which are known to influence bird distributions (Ford & Barrett, 1995; Watson *et al.*, 2009). They also represent a measure of anthropogenic disturbance (Bentley & Catteral, 1997). Each variable was measured as a percentage which was converted into a score (see Table 3.3), and the scores for the five were summed to give a total score for a point count. A habitat complexity score (HCS) was then computed for the patch by averaging these values across all point counts.

Table 3.3 The variables used to determine the patch habitat complexity. The table was adapted from information in Newsome & Catling (1979) and Watson *et al.* (2009).

Variable	Score			
	0	1	2	3
Tree canopy cover (%)	0	<30	30-70	>70
Shrub layer cover (%)	0	<30	30-70	>70
Herb layer cover	<0.5m (sparse)	>0.5m (sparse)	<0.5m (dense)	>0.5m (dense)
Ground/debris cover (%; e.g. logs, rocks)	0	<30	30-70	>70
Leaf litter cover	0	<30	30-70	>70

3.7 Bird trait data collection

3.7.1 Continuous traits

For present purposes, a trait was defined as a measurable aspect of an organism that determines the organism's interaction with the environment in some way (Flynn *et al.* 2009). As I was interested in birds in forested habitat islands, I collected data on traits that are well known to be related to the functional roles of birds in forest fragments (i.e. resource acquisition, locomotion, foraging, dispersal ability, and body size; Miles *et al.,* 1987; Henle *et al.,* 2004; Sutherland *et al.,* 2004; Grant & Grant, 2006; Tscharntke *et al.,* 2008; Claramunt *et al.,* 2012). The actual measurement of the continuous traits was primarily undertaken by a colleague (Catherine Sheard), although I measured a number of species. In total, I collected data on eight continuous traits and one categorical trait. I focused primarily on continuous traits as these have been argued to produce more accurate representations of species' functional roles in ecosystems (e.g. McGill *et al.,* 2006). The continuous traits measured were bill length (from the nares to the tip, and also a second measure from the first feathering on the culmen to the tip), bill width and depth at nares, tarsus length, Kipp's Distance (distance from the longest primary to the first secondary; Kipp, 1959), total wing chord length, and tail length (Sutherland *et al.,* 2004). The majority of trait measurements were obtained from skins in the ornithology collections at the Museum of Natural History, Tring, UK. All measurements were taken according to standardized protocols (Tobias *et al.,* 2014) to homogenize measuring techniques, and all data were both spot-checked for accuracy and later examined for outliers. Any specimen that was the sole representation of its species was measured twice. Whenever possible, for each species at least two adult males and two adult females in good condition were selected for measurement and juveniles, damaged specimens, and potential hybrids were avoided. I then averaged the data from the male and female specimens to create a species average for each morphological trait. These individual traits

were then condensed into a smaller number of trait indices using PCA analysis; the full method is outlined in Chapter 9.

3.7.2 Forest specialism

A categorical trait was measured which represented a binary code defining whether each species was a habitat (forest) specialist or habitat generalist. A species was considered a forest specialist if forest was deemed to be of major importance to the life history of the species. To classify species into specialists and generalists I used a combination of books, articles, online reference databases and expert opinion. The books were largely accessed through the Alexander Library of Ornithology, University of Oxford. Where necessary, articles were obtained through a literature search using combinations of the following keywords: "birds", "habitat specialisation", "forest specialists", "habitat specificity" and a given region, for example "South America". A number of studies (e.g. Anciães & Marini, 2000; Salisbury *et al.,* 2012) had already classified a large number of species based on habitat specificity (e.g. Salisbury *et al.,* 2012, classified 739 species in the Amazon), which provided me with a resource with which to compare with my own classifications. Finally, I also consulted a number of ornithological experts (e.g. Dr Andy Gosler; University of Oxford) for additional opinions on particular species.

3.8 Azorean arthropod data

For my empirical chapters/papers on species abundance distributions I used a well specified dataset of arthropods in forest fragments in the Azores. These data are part of a long-term ecological study conducted in the Azorean archipelago: the BALA project (1999-2012; Biodiversity of Arthropods from the Laurisilva of the Azores; see Borges *et al.*, 2005; Ribeiro *et al.,* 2005). I was not involved in the collection of these data, but as I used them in a number of analyses I summarize the dataset here.

As part of this study a total of 100 transects (150 m x 5 m) were randomly placed within 18 fragments of protected native forest in seven Azorean islands (e.g. see Fig. 3.8). In each transect, arthropods from the soil (mainly epigean) and herbaceous vegetation were surveyed with a set of 30 pitfall traps (Borges *et al.*, 2005), while arthropods associated with dominant woody plant species were sampled using a beating tray, focused on three tree species (10 samples per tree; for details see Ribeiro *et al.*, 2005). In total, 6770 samples (3420 pitfall traps and 3350 beating samples) were collected. The taxonomic groups were selected to cover the majority of arthropod diversity, excluding only those groups for which taxonomic expertise was not available. All Araneae, Opilionida, Pseudoscorpionida, Myriapoda and Insecta (excluding Diptera and Hymenoptera) were assigned to species or to morphospecies through comparison with a reference collection. Most taxa were assigned additional classifications based on colonization status (termed here 'indigenous' or 'introduced') and ecological information (predator, herbivores, saprophagous). For the purposes of my thesis analyses, I followed the classifications used in the BALA dataset and classified as introduced those species known from regions outside the Macaronesian archipelagos and that are believed to have been introduced after human settlement in the Azores in the 15th century. This was generally inferred either from historical records of detected species introductions, or from the current distribution of a species being closely associated with human activity. Otherwise, I assumed the species to be indigenous (see Borges *et al.*, 2010). For roughly 16% of the morphospecies only the family or genus is known. I classified such morphospecies as indigenous or introduced based on the dominant category (i.e. introduced or indigenous) of species in the family or genus (Borges *et al.*, 2010). The majority of species (84 % of the overall species richness) belong to four taxonomic orders (Araneae, Hemiptera, Coleoptera, and Lepidoptera; more details in Gaspar *et al.*, 2008).



Figure 3.9 A Laurisilva forest fragment on the island of Terceira, Azores. Photo courtesy of Paulo Borges.

CHAPTER 4

Thresholds and the species–area relationship: a synthetic analysis of habitat island datasets

Paper published in Journal of Biogeography: Matthews, T.J., Steinbauer, M.J., Tzirkalli, E., Triantis, K.A. & Whittaker, R.J. (2014) Thresholds and the species–area relationship: a synthetic analysis of habitat island datasets. *Journal of Biogeography,* **41**, 1018-1028.

4.1 Preface

This chapter is the first empirical chapter, and examines the prevalence of thresholds in the island species–area relationship (ISAR). A variety of piecewise regression models were compared with standard linear regression models using different data transformations. It was found that thresholds were apparent in only a small number of datasets, but that methodological issues, particularly the use of data transformations, made inference problematic. A number of recommendations for future ISAR threshold studies are provided.

Authors: Thomas J. Matthews, Manuel J. Steinbauer, Elli Tzirkalli, Kostas A. Triantis, Robert J. Whittaker

The co-author contributions were as follows. T.J.M., K.A.T., M.J.S., R.J.W. conceived the ideas; E.T. and T.J.M. collected the data; T.J.M. and M.J.S. undertook the statistical analysis; and T.J.M. and R.J.W. were responsible for development of the analyses and idea and wrote the manuscript with input from M.J.S. and K.A.T.

4.2 Abstract

Aim The application of island biogeography theory in habitat fragmentation research assumes a simple relationship between species richness and fragment area. However, previous work has suggested that in some cases thresholds can be detected, at which the form of the island species–area relationship (ISAR) changes abruptly. Piecewise regression has been advocated as a suitable statistical technique to model such thresholds. Here we first provide a comparative analysis of piecewise regression models to determine the prevalence and type of thresholds in habitat island ISARs. Second, we evaluate piecewise regression as a method for locating thresholds in the ISAR, with particular emphasis on the implications of data transformation.

Location World-wide.

Methods Seventy six habitat island datasets were sourced from the literature. An information theoretic approach was employed to compare linear regression ISAR models with piecewise regression models. The models were applied to untransformed (species–area), semi-log (species–log-area) and log–log (log species–log area) data. Three types of piecewise regression models were evaluated: continuous, discontinuous and zero slope. Model performance was compared using the Akaike's information criterion. We also examined the influence on model performance of taxon, number of habitat islands, and area of smallest island.

Results Linear regression models performed best, although piecewise models were preferred in a number of cases. Cases in which no model was significant were most prevalent in untransformed space relative to the semi-log and log–log transformations. Piecewise fits were more prevalent in datasets with a larger numbers of islands.

Main conclusions Data transformation is a key part of model selection and needs to be explicitly considered, especially in terms of drawing inferences from models. Piecewise

models, even if selected as the favoured model in our analyses, were often ecologically unintelligible in relation to area alone. When detected, breakpoint values ranged over five orders of magnitude, although with one exception all were under 50 ha. Our findings highlight the limitations of using individual threshold values to inform conservation practice.

4.3 Introduction

The destruction and fragmentation of natural habitat are primary causes of terrestrial biodiversity decline (Wilcove, 1987; Primack, 2010) and constitute a major focus of research within conservation biogeography (Ladle & Whittaker, 2011). The impacts of habitat loss and fragmentation on species diversity take time to be fully realized (Halley *et al.*, 2014), leading to a reliance on the use of various forms of species–area relationship (SAR) to quantify and predict likely future levels of species loss from landscapes (but see also, e.g., Stork, 2010). The two principal variants of SAR are species accumulation curves (SACs), which record the cumulative sum of species as sampling area is increased, and island species–area relationships (ISARs), which record the number of species found versus area for each island or habitat patch (Box 4.2 in Whittaker & Fernández-Palacios, 2007). Both variants (and others) have been used in efforts to model and predict species losses arising from habitat fragmentation. However, as the majority of fragmentation studies use ISARs the empirical tests presented herein are exclusively concerned with ISARs.

ISARs are typically interpreted in terms of MacArthur & Wilson's (1967) equilibrium theory of island biogeography. However, debate continues as to the efficacy and applicability of island theory in the context of habitat islands (e.g. Laurance, 2008). Of particular concern is whether there may be evidence for nonlinearity, perhaps involving thresholds across which relationships change form, and how analytically such nonlinearity may best be dealt with (Lomolino & Weiser, 2001; Gentile & Argano, 2005; Ficetola & Denoël, 2009; Dengler, 2010).

The definition of nonlinearity in this context is itself problematic. In general, ISARs, whether fitted to 'habitat island' or 'true island' datasets, are rarely best fitted by linear models using untransformed data. Rather, partly for analytical convenience, linear fits have been obtained by log-transformation of data, with much debate as to whether linear fits are more satisfactorily obtained using 'species richness–log area' (hereafter semi-log) or 'log species richness–log area' (hereafter log–log) transformations. The implication of a *linear* regression fit using a semi-log transformation is of course that species richness declines *nonlinearly* with declining area. For example, Drinnan (2005) uses this logic to identify a threshold area below which the loss of species accelerates with further area loss. An alternative approach to identifying thresholds is to use piecewise regression models that fit two separate linear functions to upper and lower parts of the distribution of values of area (e.g. Gentile & Argano, 2005). However, when using log-transformed area and/or species richness data, it should be understood that any breakpoint identified is in essence identifying a change of form within an already nonlinear space (Burns *et al.*, 2009). Hence, any interrogation of threshold relationships using piecewise regression requires careful interpretation of the data transformation applied.

The use of piecewise regression to test for ecological thresholds has been advocated by several authors as it allows for rigorous identification of patch area thresholds, and evaluation of the associated error of such identifications (e.g. Toms & Lesperance, 2003; Ficetola & Denoël, 2009). Ficetola & Denoël (2009) sourced 26 papers from the literature and reviewed the methods used to detect thresholds in the relationship between habitat amount and species occurrence or richness. They found that only 35% of the studies examined used appropriate statistical techniques. A number of studies were found to have used generalized linear models whereby the breakpoint was determined visually. Ficetola & Denoël (2009) contend that visual identification of breakpoints is inadequate, arguing that

only techniques such as piecewise regression have the ability to estimate both the position and probability of thresholds in habitat amount.

Previously reported threshold ISAR relationships mostly take two basic forms, which we characterize herein as shallow–steep and steep–shallow, by which we refer to the change in form of the ISAR with *increasing* area. Steep–shallow relationships are those showing an initially rapid rate of increase in richness with area, until a threshold is reached beyond which a slow rate of increase applies (Fig. 4.1). Steep–shallow thresholds indicate that the effect of habitat loss becomes exacerbated below the threshold level of habitat cover. This may be due, for example, to a patch extinction rate exceeding its immigration rate (Fahrig, 2001). A certain threshold area may also relate to coinciding minimum patch area requirements for multiple species or to a point at which the effects of habitat isolation are exacerbated (Andrén, 1994). Detecting steep–shallow relationships is an important task for conservation planners (Drinnan, 2005; Huggett, 2005). Shallow–steep ISARs are those that initially show a slow rate of (or no) increase in richness with increasing area, followed by a steeper phase beyond a critical threshold. Shallow–steep thresholds include those that represent the so-called small island effect (SIE: e.g. see Lomolino & Weiser, 2001), which has previously been described for some habitat island datasets (Qie *et al.*, 2011).

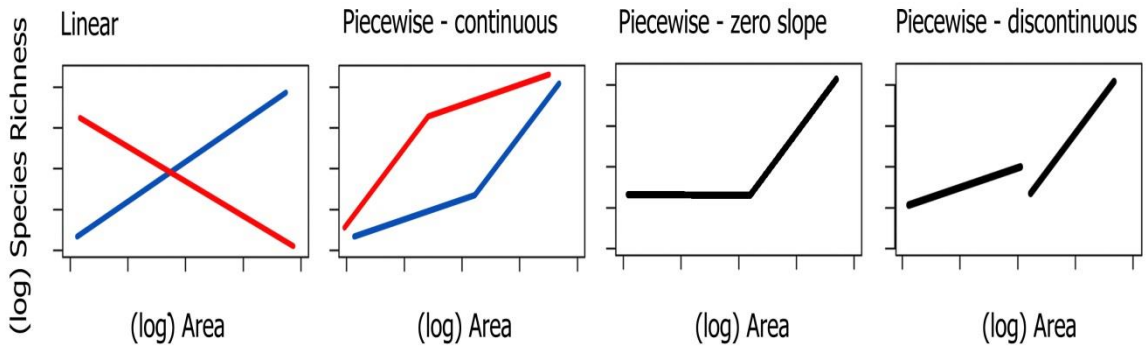


Figure 4.1 Schematic illustration of the alternative models used in the analysis, and their different forms. In the 'linear' plots the red line signifies a negative linear relationship and the blue line a positive linear relationship. In the continuous plots the blue lines signify a shallow–steep form (following the line from left to right) and the red line indicates a steep–shallow form of the model. The discontinuous plots do not represent the sole form of a discontinuous model; rather multiple patterns emerged through the analysis (see Fig. A2.1 in Appendix 2). We denote the *x*-axis as (log) Area, and the *y*-axis as (log) Species richness, but have applied the models to data that were untransformed, semi-log and log–log transformed.

Piecewise regression models incorporate the inclusion of breakpoints within linear regressions, with the breakpoints representing a point at which the form of the relationship changes, either in slope or in slope and intercept (Toms & Lesperance, 2003; Gentile & Argano, 2005). Analytically, the possible forms of these relationships are in practice more numerous than the two basic forms described above and include the following set of piecewise models: continuous, zero slope, and discontinuous (Fig. 4.1). Continuous piecewise models incorporate a change in slope at the break-point, with the condition that the regression line being modelled is continuous. Zero slope models are a subtype of the shallow–steep continuous model whereby the slope to the left of the breakpoint is zero (i.e. there is no relationship). Zero slope models are included within the model comparison as they are recommended by Lomolino & Weiser (2001) for detection of the SIE. It is important to note, however, that this methodological approach to detecting SIEs has been the subject of dispute (e.g. Triantis *et al.*, 2006; Dengler, 2009, 2010; Tjørve & Tjørve, 2011; Triantis &

Sfenthourakis, 2012). In a discontinuous piecewise model both the slope and intercept of the regression line can change at the breakpoint; that is, the regression line is not continuous. In addition to the breakpoint models, around 20 functions have been described for fitting ISARs (Triantis *et al.*, 2012). We do not attempt to determine the fit of these alternative ISAR models as our study is focused on the fit of piecewise models and their use in detecting area thresholds.

Despite existing theoretical and empirical work focusing on patch area and extinction thresholds (e.g. Drinnan, 2005; see Swift & Hannon, 2010, for a review), few previous studies have re-analysed existing datasets to explore emergent patterns (but see Ficetola & Denoël, 2009). Moreover, we are unaware of any statistical re-analysis of patch area thresholds, or for that matter any type of ecological threshold, incorporating a large number of datasets. The aim of this paper is thus twofold. First, we provide a synthetic analysis of piecewise regression models (Fig. 4.1), using an information theoretic approach (Burnham & Anderson, 2002) to determine the prevalence and type of thresholds in the ISAR of 76 habitat island datasets. Having done so, we examine the resulting model fits with reference to the distribution of island areas and other properties of the source datasets. Second, we evaluate piecewise regression as a method for locating thresholds in the ISAR, with particular emphasis on data transformation. We go on to provide methodological recommendations and comment on the implications of our findings for conservation biogeography.

4.4 Methods

4.4.1 Data collection

Three abstracting services were searched for relevant habitat loss and fragmentation studies and datasets during May–June 2007 and May–July 2011: JSTOR (1913–2003); ISI Web of Knowledge (1980–2011); and BIOSIS Biological Abstracts (1980–2003). The search

keywords were ‘habitat islands and species richness’ and ‘forest fragments and species richness’. Additionally, cross-referenced papers derived from the reference lists of sourced papers were also included.

Suitability was based on the following criteria:

1. Habitat islands were defined as discrete patches of habitat surrounded by contrasting, ‘hostile’ matrix habitat.

2. The areas and species richness of each habitat island were listed.

3. The number of habitat islands within each dataset was at least eight, as this was necessary so that all the available piecewise models could be applied (below).

4. Data from any particular study did not overlap with those from any other study accepted for analysis (data for different taxa within the same study system were allowed).

For each dataset we recorded the geographical location, the taxon studied, the habitat island type (forested or non-forested), and the range of species richness and island sizes (see Appendix 2).

4.4.2 Data analysis

First, we developed three versions of the datasets: untransformed (*S* versus *A*), semi-log (*S* versus log*A*) and log–log (log*S* versus log*A*) transformed, where *S* is species richness and *A* is area. Second, we fitted linear models to all three versions of each dataset alongside a set of piecewise models: continuous piecewise, zero slope continuous and discontinuous piecewise regression models. As log(0) is non-defined, a log(x + 0.1) was applied for the transformation

of species richness in the case of the log–log transformation. All analyses were performed using R 3.0.0 (R Development Core Team, 2013).

In fitting the piecewise models, the optimal breakpoints for each continuous and discontinuous piecewise model and dataset were selected by testing 1000 evenly distributed points within the range of the data. The optimal breakpoint leading to the best-fitting model was selected from this set by a log-likelihood optimization separately for each case and model type. A benefit of this methodology is that the existence of thresholds (breakpoints) in the data is not assumed a priori; this being a criticism often levelled at such studies (see Gentile & Argano, 2005).

All models were implemented and compared as linear regressions. For each of the three versions of each dataset (i.e. untransformed, semi-log or log–log) the performance of each model was compared using Akaike's information criterion corrected for small sample size (AIC_c; Burnham & Anderson, 2002; code from library 'AICcmodavg' version 1.20; Mazerolle, 2011):

$$AIC_c = -2 \times \text{logLik(model)} + 2 \times k \ [n/(n - k - 1)],$$

where n is the number of observations and k the number of model parameters. The selection of a single best model (i.e. the statistically best-fitting descriptive model) was based on there being a difference in AIC_c value of at least two between the best and a respective weaker model (Burnham & Anderson, 2002).

In addition, because we were concerned that the initial optimization of the breakpoints could be deemed to constitute 'hidden parameters', we also performed a sensitivity analysis in which we ran the model selection after accounting for the initial optimization of the breakpoints and increase in model complexity by increasing the number of model parameters to $k = 4$ for the piecewise continuous model with zero slope, $k = 5$ for the piecewise continuous model, and $k = 6$ for the piecewise discontinuous model, when calculating AIC_c.

As the calculation of AIC_c includes a division by $n - (k - 1)$ (with n being the number of data points and k the number of model parameters), to avoid a negative denominator the minimum number of data points was set to eight. In addition to the aforementioned linear and piecewise models, we also implemented a model with a zero slope regression line at the data mean as our simplest model (corresponding to no species–area relationship). Cases in which none of the more complex models were favoured above this 'null model' based on our AIC_c criterion were treated as 'non-significant'. As an additional criterion, a standard *F*-test was used to check the significance of the species–area relationship for each model. Only models showing a significant species–area relationship were compared based on AIC_c. In a number of datasets no model showed a significant relationship. These cases were also labelled as 'non-significant'. A direct comparison of models across different versions of the dataset is not possible based on AIC_c values as the model types have different dependent variables [log(S + 0.1) and S].

In a further step all analyses were re-run following the application of a Cook's distance criterion (Cook & Weisberg, 1982). This was done to remove models which were unduly influenced by a single data point. Many habitat island datasets include relatively few islands (Appendix 2) and thus outliers are a significant problem that needs to be considered. For each version of the dataset, models in which any data points had a Cook's distance greater than one were removed from the model comparison for that version of the dataset (Cook & Weisberg, 1982). We summarize the initial results in Appendix 2 and present the results following application of the Cook's distance criterion in the main Results (below).

Plots of each dataset were visually checked to allow us to categorize linear models as either positive or negative and continuous piecewise models as either shallow–steep or steep–shallow forms (as Fig. 4.1). In addition, we took note of visually peculiar relationships to inform subsequent interpretation.

4.4.3 The effects of the dataset characteristics

Model performance was compared between taxonomic groups, habitat type and the various dataset characteristics recorded (number of habitat islands, range of habitat island areas, and area of smallest and largest habitat island) using the Student's *t*-test, ANOVA and Tukey's honestly significant difference (HSD) post-hoc test. To determine the effect of habitat type, datasets were separated into those comprising forested habitat islands and those comprising non-forested habitat islands and model performance compared using a test of proportions (prop.test in the basic STATS package in R). After the initial analysis the dataset selection criteria (above) were iteratively altered to progressively remove datasets with the fewest habitat islands, increasing in units of one from our pre-set minimum of eight, up to 32 habitat islands. After each iteration the analyses were re-run to determine the effect of sample size on the results (following Triantis *et al.*, 2012).

4.5 Results

Over 500 articles were reviewed, from which we retrieved 76 datasets meeting the above criteria from 70 separate sources (Appendix 2). There were 53 forest habitat island datasets (70%) and 23 non-forest systems (30%); and there were 14 (18%) invertebrate, 39 (51%) vertebrate, and 23 (31%) plant datasets. The vertebrate datasets comprised 22 bird, 15 mammal and two reptile case studies.

4.5.1 Model performance

The application of Cook's distance criterion resulted in the best model changing for four of the untransformed datasets, four of the semi-log transformed datasets, and six of the log–log transformed datasets (see Table A2.1 in Appendix 2 for overall outcome, which remains largely unchanged). Whilst this process removed a number of visually idiosyncratic, ecological unintelligible models, inspection of the plots revealed that several remained that

defy ecological interpretation in terms of isolate area ($n \approx 8$; Fig. A2.1 in Appendix 2): these latter cases were nonetheless retained in the analysis (Table 4.1).

Unsurprisingly, the best model type varied depending on the data transformation undertaken prior to model fitting (Table 4.1). Despite the range in form of the models used, cases where no significant model was generated accounted for 51% of the untransformed datasets, 32% of the semi-log datasets and 37% of the log–log transformations, respectively. Positive linear fits were the commonest fit and were found in respectively 18%, 32% and 32% of cases. The second best model for untransformed datasets and for the log–log datasets was the subset of piecewise continuous models visually classified as steep–shallow, with 22% and 16%, respectively, while the zero slope piecewise model was the second commonest best model in semi-log space, being selected for 14% of cases. The breakpoint values of piecewise models selected as the best were found to vary over five orders of magnitude (0.3–23,000 ha), although with one exception all were under 50 ha.

The sensitivity analysis in which we re-ran the model selection after increasing the number of parameters to take explicit account of the initial optimization of the breakpoints, resulted in a small number of changes to the models chosen as best (compare Table 4.1 with Table A2.2 in Appendix 2). In particular, we found greater support for linear models relative to piecewise models. However, qualitatively our results did not change, and from herein we only discuss results based on the standard AIC_c calculation.

Table 4.1 The best fit island species–area relationship (ISAR) model for 76 habitat island datasets, tallying the best fit models separately for each data transformation. The best model for each version of the dataset (i.e. untransformed, semi-log and log–log) was chosen based on Akaike's information criterion corrected for small sample size ($\Delta AIC_c > 2$). These results were tallied following the application of Cook's distance criterion: for results prior to this step, see Appendix 2.

	Number of datasets (Percentages in parentheses)		
Model Type	Untransformed	Semi-Log	Log-Log
Linear (positive relationship)	14 (18%)	24 (32%)	24 (32%)
Linear (negative relationship)	1 (1%)	0	0
Piecewise continuous (steep-shallow)	17 (22%)	5 (7%)	12 (16%)
Piecewise continuous (shallow-steep)	0	6 (8%)	0
Piecewise Zero Slope (i.e. flat-steep)	1 (1%)	11 (14%)	5 (7%)
Piecewise Discontinuous	4 (5%)	6 (8%)	7 (9%)
Non-Significant	39 (51%)	24 (32%)	28 (37%)

4.5.2 Dataset descriptors

The proportion of cases where at least one model was significant in at least one version of the datasets was lower for invertebrate datasets (57% of cases), than for vertebrates (74%) and plants (74%) (detailed results not shown). The small sample sizes involved prohibit formal statistical testing of between-taxa differences. However, the breakdown of results by taxon shows that setting aside linear models, piecewise continuous models emerged as the best most often from vertebrate datasets, whereas discontinuous piecewise models were more often found for plant datasets (Table A2.3 in Appendix 2). There was no significant difference in the proportion of piecewise models selected as the best between forested and non-forested habitat islands ($P = 0.32$; prop.test).

Grouping the models into piecewise and non-piecewise it was found that the mean number of islands in datasets in which piecewise models performed best for at least one version of the dataset (mean ± SD = 26 ± 12, n = 38) was greater than for the remaining datasets (mean ± SD = 19 ± 12, n = 38), although this difference was non-significant (t = 2.85, P = 0.1; Student's t-test). Datasets in which no models were significant generally had fewer islands than datasets with a significant ISAR; although the differences between the number of islands in datasets with no significant model and datasets best modelled by a linear or piecewise model was not always significant according to Tukey's HSD post-hoc test (Fig. 4.2a–c). Versions of datasets in which a discontinuous model was selected as best mostly had more islands than those in which other models were selected as best. However, again these differences were not always significant according to Tukey's HSD test (Fig. 4.2a–c). The size of the smallest island was smaller in versions of the datasets in which discontinuous models performed best, in comparison to those in which a linear model was best, or in which no model was significant (Fig. 4.2d–f). However, these differences were not always significant and, in general, the minimum island size in a dataset does not clearly discriminate between piecewise continuous (shallow–steep and steep–shallow forms) and linear models. There was no significant difference in the proportion of piecewise models selected as best based on either the area of the largest island in the dataset, or the overall range of habitat island areas in the study (results not presented). As some of the foregoing comparisons involve very few datasets, caution is necessary in interpreting these findings.

Progressively removing datasets with fewer habitat islands led in general to a larger fraction of datasets being best approximated by piecewise models and a reduction in the number of non-significant models (Fig. 4.3).

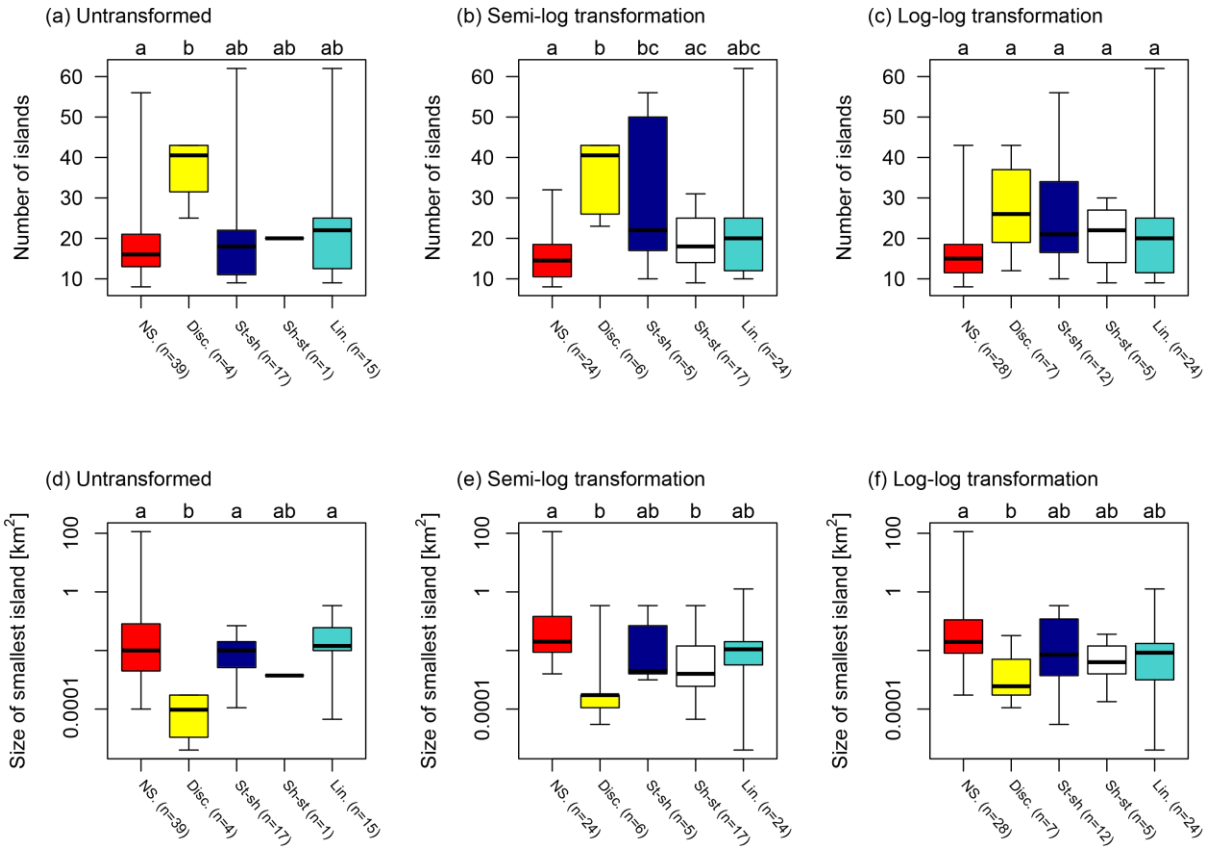


Figure 4.2 Box plots displaying the number of habitat islands (top row), and area of the smallest habitat island (bottom row), for each model type across the three data transformations. Significant differences between models are displayed as different lowercase letters above the box plots. Models that do not significantly differ in the dependent variable have the same lowercase letters. NS refers to datasets where no model provided a significant fit, 'Disc' = datasets in which a discontinuous model was selected as best, 'st-sh' = steep–shallow, 'sh-st' = shallow–steep, and 'lin' = a linear model. Shallow–steep refers to both shallow–steep continuous models and the zero-slope piecewise models. Note that small '*n*' values in particular categories limits the possibility to verify differences significantly. Significance was assessed by ANOVA followed by Tukey's HSD (honestly significant difference) post hoc test ($P = 0.05$).

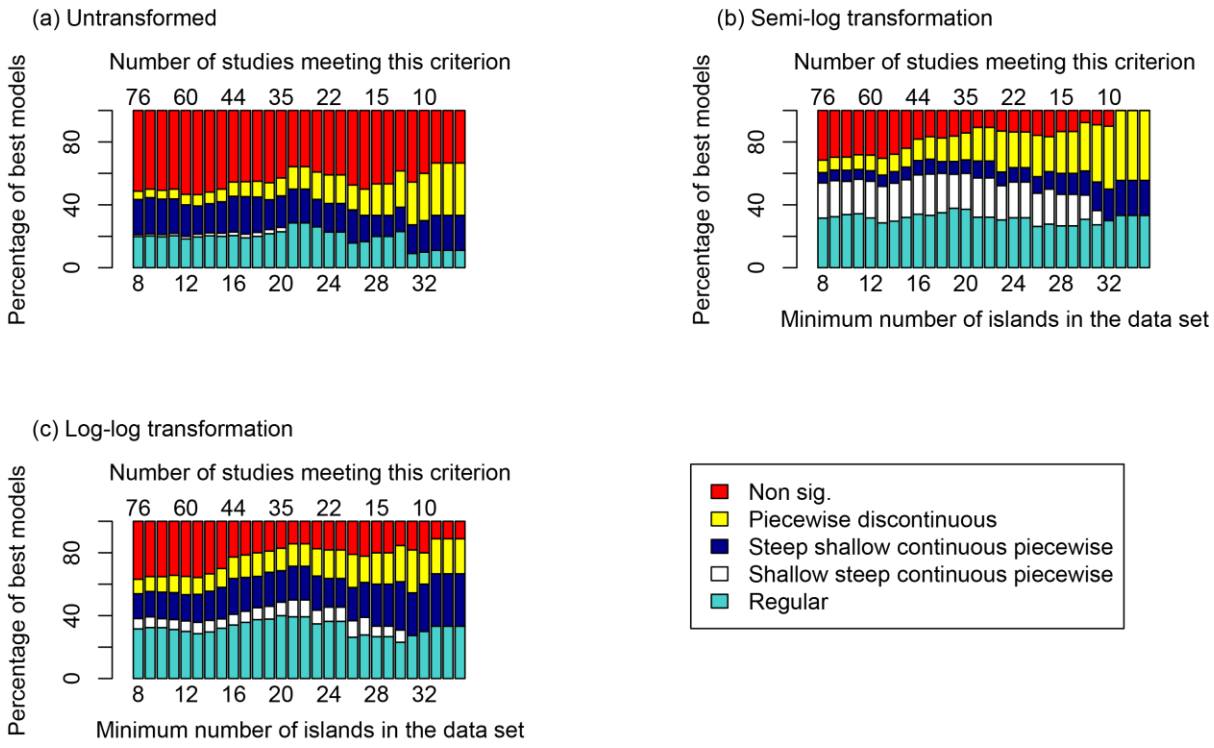


Figure 4.3 Percentage of models identified to be the best depending on the minimum number of habitat islands (sample size) of a dataset selected, for the three versions of each dataset (untransformed, semi-log and log–log). Note that eight represents the minimum number of habitat islands allowed by the original selection criteria for analysis. Non-sig = non-significant; regular = linear fit.

4.6 Discussion

4.6.1 Data transformation

To determine the prevalence of thresholds in the ISAR of habitat islands (n = 76 datasets), statistical techniques incorporating a breakpoint in the linear regression were compared with simple linear regression species–area models. The models were applied to untransformed, semi-log and log–log transformed data. We found that linear regression models generally performed best, with a small number of different versions of the datasets best fit by piecewise regression models.

Our analyses serve to highlight the danger of taking an a priori decision on data transformation without careful reflection on study aims. Such a strategy may obscure

significant variations in the form of the ISAR across data transformations and lead to erroneous (or incomplete) inferences surrounding the shape of the ISAR, and, ultimately, the identification of thresholds. It is important to be aware of these pitfalls and to ensure that the analytical strategy adopted is appropriate for the purpose of the particular study because it is possible for two quite different models to provide good fits to the same dataset. Figure 4.4 illustrates this for a dataset drawn from Watson (2003), a study of naturally fragmented montane habitat islands in Mexico. Whereas the best model in untransformed space is a piecewise continuous model of the steep–shallow form, the best in semi-log space is a zero slope piecewise model. Therefore, depending on the transformation used one could infer that there is a habitat amount threshold below which richness precipitously declines, or evidence of a traditional small island effect (*sensu* Lomolino & Weiser, 2001). With regard to the zero slope pattern found for the semi-log analysis, the data points responsible for the initial lack of area trend each had very low (or no) species (Fig. 4.4). These data points almost overlay each other in the graph of the untransformed data, while the semi-log transformation in effect expands this section of space to generate the zero slope pattern.

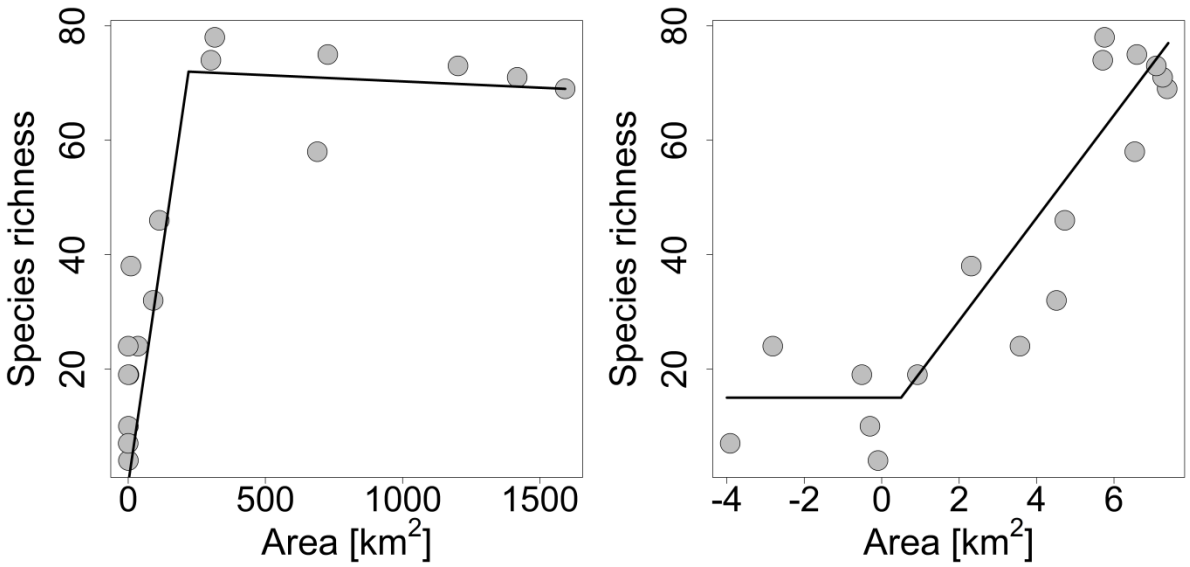


Figure 4.4 Variation in the best model selection across two versions of the same dataset. Left, the data plotted in untransformed space in which the best model is a continuous (steep–shallow) piecewise model. Right, the same data plot plotted in semi-log space, in which the best model is a zero slope piecewise model. Data are from Watson (2003): a study on birds in a fragmented landscape, Mexico.

The dataset of Kitchener *et al.* (1980) also provides a useful illustration of the change in model selection across the different versions of the dataset (Fig. 4.5). Visual examination of the untransformed plot suggests a threshold but the best model is actually a simple linear (positive) model. However, the best model in semi-log space is also a simple linear regression and thus, using the method of Drinnan (2005) this would be equivalent to the identification of a threshold in untransformed space; while for the record a steep–shallow piecewise continuous model was preferred in log–log space.

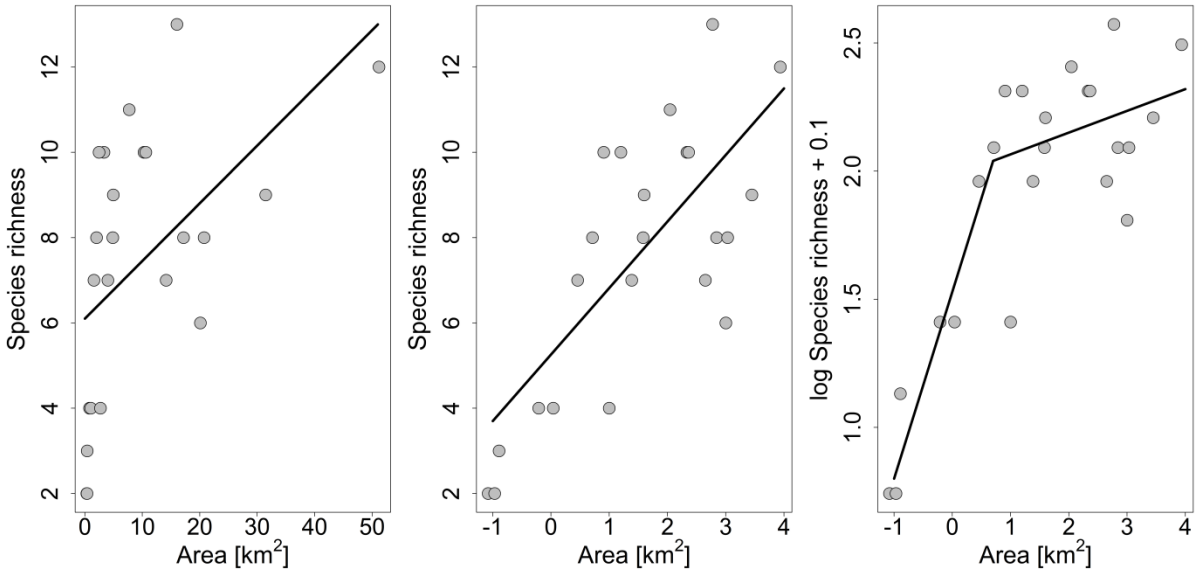


Figure 4.5 Variation in the best model selection across three versions of the same dataset. Left to right: untransformed, semi-log, log–log. The best model for the untransformed and semi-log transformation was simple linear (positive), and for the log–log transformation, the best was piecewise continuous (steep–shallow). Data are from Kitchener *et al.* (1980): a study on mammals in Australia.

Hence the question arises as to what a linear (or alternatively a piecewise) relationship actually signifies when using logged variables. For instance, a linear relationship in semi-log space equates to a curvilinear relationship using the untransformed data. Thus, given that 34% of datasets are best modelled by a linear model in semi-log space, an underlying curvilinear form can be described for about half of the datasets that we are able to fit ISARs to. However, our results indicate that the higher proportion of linear models that performed best in semi-log and log–log space does not correspond to an equally high proportion of piecewise models in untransformed space (Table 4.1). Rather, datasets in which no model provided a significant fit were in the majority in untransformed space. The likelihood of detecting thresholds and what form they take is thus to a considerable degree dependent on the data transformation applied before analysis: an obvious point mathematically but one that

is frequently given scant attention in ecological threshold analyses (but see Wilson, 2007; Burns *et al.*, 2009).

4.6.2 Interpreting discontinuous models

Scrutiny of plots of all 'best' models revealed that several of the piecewise models are difficult to interpret ecologically, particularly in the case of discontinuous piecewise models (see Fig. A2.1 in Appendix 2). While it is conceivable that discontinuous relationships may exist in nature (see Maron *et al.*, 2012), their detection in a few datasets most likely signifies important roles for confounding variables that were not included in our models, such as isolation or matrix effects (e.g. see Crowe, 1979; Levenson, 1981). For instance, the best model in log–log space for the Crowe (1979) dataset was a discontinuous piecewise model. Crowe (1979) tested various predictions of the equilibrium theory of island biogeography (MacArthur & Wilson, 1967) using plants in urban habitat patches of varying age in the USA. It was found that richness in older patches (over 40 months in age) had no relationship with patch age and did exhibit a significant ISAR, whereas young patches (under 40 months) exhibited a significant species–age relationship and no ISAR. Crowe (1979) suggested that older patches could be considered at equilibrium whereas younger patches were still 'acquiring' species through colonisation. This suggests that the selection herein of a discontinuous piecewise model is uninformative of the effect of area, but instead reflects a confounding variable.

The issue of confounding variables is not restricted to discontinuous models. For example, Watson (2003) identifies the same steep–shallow threshold and breakpoint in untransformed space that we have identified. But he states that there is a high degree of collinearity within his independent variables and thus this pattern cannot be solely attributed to area but may be a product of, for example, elevation, patch configuration, or number of microhabitats. We have not conducted formal analyses of such potentially confounding

system variables because there may be many relevant factors and their availability is limited and inconsistent within the source papers (cf. Whittaker, 2010).

4.6.3 Ecological interpretation

It is intriguing that piecewise continuous models emerged as best more often for the vertebrate datasets, with the steep–shallow form commonest. This observation is mirrored by the fact that piecewise discontinuous models emerged as best for 3 (13%), 5 (22%), and 4 (17%) of the untransformed, semi-log transformed and log–log transformed plant datasets respectively, but for none of the untransformed versions of the vertebrate datasets, and only 1 (3%) semi-log transformed and 3 (8%) log–log transformed vertebrate datasets (Table A2.3 in Appendix 2). There were less obvious emergent patterns for the invertebrate datasets, although it is apparent that invertebrate datasets had a higher proportion of cases in which no model provided a significant fit compared to the vertebrate and plant datasets, for the semi-log and log–log versions of the datasets. While the small sample sizes make it dangerous to speculate on these apparent differences, these results suggest differences in the responses of the three taxa to habitat loss that merit further investigation.

4.6.4 Conservation and methodological implications

Patch area thresholds have been posited as an important component of protected area network design at the landscape scale (Drinnan, 2005; Huggett, 2005). As such, we offer the following observations with respect to future research on the topic.

Piecewise regression has been advocated as an approach to identifying thresholds on the basis that it is more objective than methods such as visual inspection of breakpoints (Ficetola & Denoël, 2009). However, we have shown above that this approach requires careful consideration with regard to the choice of threshold model and data transformation (see also Connor & McCoy, 1979; Burns *et al.*, 2009; Dengler, 2009). This is clearly illustrated for the Watson (2003) dataset in which a single dataset reveals two seemingly

opposing patterns depending on the transformation (Fig. 4.4). It could be argued that both these patterns provide partial truths about the dataset. However, we have also noted that there are confounding variables that may account for the features of the ISAR (see Watson, 2003). Thus, where possible future empirical studies of patch area thresholds should consider other potentially important variables, such as isolation, properties of the matrix and patch age.

We found a small amount of support for the zero slope model: 11 datasets (14%) were best categorized by the zero slope model in semi-log space, and five (7%) in log–log space, and therefore exhibited what has traditionally been described as a SIE (*sensu* Lomolino & Weiser, 2001; e.g. see the right hand side plot of Fig. 4.4). All that can be said is that the SIE pattern may be identified statistically in a low proportion of habitat island systems (see also Qie *et al.*, 2011), although it should also be noted there are other methods of detecting SIE patterns (see Triantis *et al.*, 2006) that were not explored in this paper.

In terms of specific recommendations regarding data transformation in the application of piecewise models one stratagem is to fit ISARs to the untransformed data, semi-log transformed data, and log–log transformed data and consider all three results to generate a conclusion. However, in doing this it is obviously necessary to interpret a particular model in the context of the transformation used (Wilson, 2007). Our results also suggest that the choice of transformation may influence the detection of particular thresholds. For instance, if the aim of a study is to detect classical steep–shallow thresholds in a set of habitat fragments for conservation management purposes it is not advisable to rely upon piecewise regression of a semi-log transformed dataset, for the reasons outline above and because steep–shallow thresholds were detected very infrequently in semi-log transformed data (see Table 4.1).

In addition to the complexity surrounding the use of data transformation we identify three further points where future studies may advance statistical methodologies in the search for patch area thresholds. First, use of the log–log transformation to model ISARs generally

involves the use of log(x+ ε) to avoid the problem of log(0) being non-finite. The choice of ε is often arbitrary and can affect the shape of the resulting relationship (Millar *et al.*, 2011), although this not as problematic for ISAR statistics in comparison to other data. Thus, for future studies of this nature it may be preferable to avoid transformations and apply nonlinear methodologies when fitting nonlinear relationships (Steinbauer *et al.*, 2012).

Second, a number of 'best' models were changed through our use of the Cook's distance criterion (Table A2.1 in Appendix 2). Despite all its advantages, the AIC_c framework was not developed to correct for problems associated with extreme values, and regression analyses, as with other statistical methods, need to be critically evaluated in this respect. The influence of data extremes is usually (if at all) checked using visual inspection of model residuals, which can be especially difficult in cases of small datasets, which are common in island biogeographical analyses. As such, we suggest the use of a more rigorous statistical test to highlight potential outliers, such as Cook's distance.

Third, piecewise regression may occasionally lead to ecologically uninterpretable models being selected as the 'best' when used within a model comparison framework; this was particularly the case for discontinuous models. As such, we believe a necessary stage of the process is to use contextual data and graphical display to identify instances in which the best model may be the result of a statistical or experimental design artefact as opposed to any biological mechanism connected with isolate area.

Our results do not provide the hoped for generic guidance on patch area thresholds but do highlight the importance in future research of considering the effects of data transformations and of sampling a large number of patches and a wide range of patch areas, in order to permit meaningful tests for such thresholds. Notwithstanding the methodological implications outlined above, it is perhaps noteworthy that the 17 datasets which displayed a significant piecewise continuous (steep–shallow) relationship in untransformed space (thus

bypassing the data transformation quandary) contained breakpoint values which varied over five orders of magnitude of area. Thus the search for a 'magic threshold value' to guide land management and conservation of multiple taxa (e.g. Andrén, 1994) may be misguided (Fahirg, 2001; Betts *et al.*, 2007; Montoya *et al.*, 2010). However, it should be stressed that all these breakpoints, with the exception of the Watson (2003) system, were under 0.5 km^2 (50 ha), highlighting the ecological precariousness associated with very small fragments.

In sum, it is problematic to determine the prevalence of thresholds in the ISAR of habitat islands. While various piecewise models can be fitted and when grouped may be deemed statistically 'best' in about the same number of cases as can simple linear models, it is worth noting the following: first, that we have tried fitting four different types of piecewise models, thus giving every chance for a piecewise model to emerge as best; second, the emergence of a piecewise model as best does not necessarily mean that the resulting model will be ecologically intelligible; and third, that initial choice of data transformation is a key determinant of the outcome of model fitting. Hence, in the search for thresholds in the ISAR, we call on authors to be explicit as to which data transformations they use at each stage of analysis, and what any particular transformation may mean in terms of inferences based on these analyses. We caution that the use of piecewise regression within a single dataset, or within a meta-analysis framework, could lead to erroneous inferences based on model performance criterion alone and, in particular, more thought needs to be given to the type of data transformation(s) employed.

CHAPTER 5

On the form of species–area relationships in habitat islands and true islands

Paper accepted in Global Ecology & Biogeography (November 2014)

5.1 Preface

This chapter is the second empirical chapter, and explores the form of the island species–area relationship in a set of 207 habitat island datasets. An information theoretic approach was used to compare the fit of 20 competing ISAR models to the 207 datasets. Linear modelling was used to determine the effect of various characteristics on the parameters of the power model. It was found that the power model was the overall best model of those compared, and variation in the parameters of the power model could be partly explained by a number of different variables.

Authors: Thomas J. Matthews, François Guilhaumon, Kostas A. Triantis, Michael K. Borregaard & Robert J. Whittaker

The co-author contributions were as follows: T.J.M., R.J.W. and K.A.T. conceived of and designed the study, T.J.M. collected the data, T.J.M. and F.G. analysed the data, all authors contributed to the interpretation, and T.J.M. led the writing, with input from all other authors.

5.2 Abstract

Aim We undertook the largest yet comparative study of the form of the island species–area relationship (ISAR) using 207 habitat island datasets and 601 true island datasets. We also undertook analyses of (a) the factors influencing *z* and *c* values of the power (log–log) model, and (b) how *z* and *c* vary between different island types.

Locations Global

Methods We used an information theoretic approach to compare the fit of 20 ISAR models to 207 habitat island datasets. Model performance was ranked according to pre-set criteria, including metrics of generality and efficiency. We also fitted the power (log–log) model to each dataset and analysed variation in parameter estimates and model fits as a function of key dataset characteristics using linear models and constrained analysis of principal coordinates.

Results The power (non-linear) model provided the best fit to the most datasets, and was the highest ranked model overall. In general, the more complex models performed badly. Average *z* values were significantly lower for habitat island datasets than for true islands, and were higher for mountaintop and urban habitat islands than for other habitat island types. Average *c* values were significantly lower for oceanic, and significantly higher for inland water-body islands, than for habitat islands. Values of *z* and *c* were related to dataset characteristics including the ratio of the largest to smallest island, and the maximum and minimum richness values in a dataset.

Main conclusions Our multi-model comparisons demonstrated the non-linear implementation of the power model to be the best overall model and thus to be a sensible choice for general use. As the *z* value of the log–log power model varied in relation to ecological and geographical properties of the study systems, caution should be employed when using canonical values for applied purposes.

5.3 Introduction

The species–area relationship (SAR), i.e. the general increase in the number of species recorded with increasing sampling area, is one of the fundamental patterns in biogeography (MacArthur & Wilson, 1967; Rosenzweig, 1995; Whittaker & Fernández-Palacios, 2007). A number of hypotheses have been put forward to explain SAR phenomena, including the Equilibrium Theory of Island Biogeography (ETIB), the habitat diversity hypothesis, passive sampling from a regional species pool, and increasing rate of speciation with increasing area (e.g. MacArthur & Wilson, 1967; Losos & Schluter, 2000; Triantis *et al*., 2003; Whittaker & Fernández-Palacios, 2007).

For present purposes, we focus purely on island species–area relationships (ISARs; Whittaker & Matthews, 2014). The term ISAR refers to Scheiner's (2003) type IV curve, wherein the number of species sampled within each of a set of isolates is plotted as a function of isolate area. It is also necessary to distinguish between different types of islands. Herein we define true islands as geographical islands within a matrix of water, i.e. oceanic islands, continental-shelf islands and inland water-body islands, whereas habitat islands are isolates of natural habitat (including lakes) surrounded by a contrasting non-water matrix type (see Whittaker & Fernández-Palacios, 2007).

The ISAR has broad applicability in ecology and conservation and has been used as the basis for protected area design, predicting species extinctions resulting from the loss of native habitat, and to estimate regional diversity from smaller-scale sample data (e.g. Dengler, 2009; Smith, 2010; Halley *et al.*, 2013; Gerstner *et al.*, 2014). However, despite this applied usage in terrestrial systems, most ISAR comparison studies have focused on true islands. As such, the development of conservation theory regarding habitat islands has relied too heavily on analyses of true island datasets; with insufficient attention to specific habitat island ISAR patterns (cf. Laurance, 2008; Mendenhall *et al.*, 2014a). For example, Triantis *et*

al. (2012) provided a synthetic analysis of ISARs involving comparisons of fit for 20 ISAR models to data from 601 true island datasets. No comparable analysis has been conducted using solely habitat island datasets, and as Sala *et al.* (2005, p.380; see also Halley *et al.*, 2013) state, the "precise shape of the relationship" in terrestrial systems is unknown. The comparability of ISAR patterns between true islands and habitat islands, and the application of island theory to habitat islands, are thus key themes within both conservation and countryside biogeography (Daily *et al.*, 2003; Pereira & Daily, 2006; Koh *et al.*, 2010; Matthews *et al.*, 2014a,b; Mendenhall *et al.*, 2014a, b).

The past decade has seen a rise in the number of studies examining the form of the ISAR (e.g. Dengler, 2009; Tjørve, 2009; Williams *et al.*, 2009; Triantis *et al.*, 2012). Over 20 functions have been proposed (e.g. Table 5.1), of which the power function remains the most commonly employed. Underpinning this research endeavour is a growing acceptance that the choice of function is an integral component of the applications of the ISAR (Guilhaumon *et al.*, 2008; Benchimol & Peres, 2013; Halley *et al.*, 2013). For example, linear, convex and sigmoidal models will result in very different estimates of richness hotspots when used in comparative analyses (Guilhaumon *et al.*, 2008), while the form of ISAR fitted is key to predictions of the number of extinctions resulting from habitat loss (Halley *et al.*, 2013). The growing popularity of multi-model inference methods in ISAR research (e.g. Guilhaumon *et al.*, 2008, 2010; Scheiner *et al.*, 2010; Benchimol & Peres, 2013; Matthews *et al.*, 2014a), is evidence of the increasing realisation that ISAR form matters.

Notwithstanding the recent focus on ISAR form, most applied habitat island ISAR studies employ the power model in combination with a set/pre-defined *z* value, generally Preston's (1962) canonical value of 0.25 (e.g. van Vuuren *et al.*, 2006; see examples listed in Kitzes & Harte, 2014). However, previous empirical work suggests a wide range of values can occur (see Whittaker & Fernández-Palacios, 2007). In particular, it is expected that

habitat island ISARs, being less effectively isolated, should have lower *z* values than true islands. This proposition has rarely been tested using multiple datasets and our understanding of variation in ISAR parameters remains hazy.

Few studies have explored ISAR patterns in a large number of habitat island studies. Watling & Donnelly (2006) conducted an analysis of several habitat island datasets, but fitted just a single ISAR model. Similarly, the meta-analysis of Drakare *et al.* (2006) only considered two ISAR models, and is confounded by the inclusion of nested SARs. Hence our paper has three aims. First, we use an information theoretic approach to compare the fit of 20 ISAR models to 207 habitat island datasets. Second, we test for variation in parameter estimates and model fits as a function of various dataset characteristics. Third, we compare our findings with those for true island datasets (n = 601; Triantis *et al.*, 2012), to provide the most comprehensive comparison of ISARs in the two island types to date.

5.4 Methods

5.4.1 Habitat island data collection

Between May 2010 and August 2013 we searched within JSTOR (1913–2003), ISI Web of Knowledge (1980–2013) and BIOSIS Biological Abstracts (1980–2003) using the keywords 'species richness', 'fragments', and 'habitat islands', in different combinations. Certain datasets were obtained from the authors of the source papers, whilst others were supplemented with additional data from the source paper authors. Following Matthews *et al.* (2014b), our criteria for selecting datasets were: (i) habitat islands constituted discrete patches of habitat surrounded by contrasting habitat (we also included a small number of datasets consisting of protected areas in which the contrast between the islands and the intervening matrix was not so pronounced); (ii) there were at least four habitat islands (as Triantis *et al*., 2012); (iii) the area and number of species of each habitat island were known;

(iv) data did not overlap with those from any other study already accepted for analysis (data for different taxa within the same study system *were* accepted).

We also included two datasets of birds in 40 habitat islands in fragmented landscapes in North France and South Spain, collected by the first author (see Appendix 3 for details). For each dataset we recorded: the latitudinal midpoint of the habitat patches (*Lat*; for certain datasets this was an estimate as precise data were not presented in the source paper), taxon (*Tax*), number of islands (*Ni*), area of smallest (*Amin*) and largest island (*Amax*) and the ratio between them (i.e. *Amax*/*Amin*; *Ascale*), minimum (*Smin*) and maximum (*Smax*) species richness values and the ratio between them (*Sscale*). We categorised habitat island types (*Typ*) as forest islands, grassland fragments, mountain-top islands (*sensu* Brown, 1971), urban fragments, and other (e.g. gravel pits, protected areas). Forest island datasets were defined as those in which the main vegetation type was forest, and the surrounding matrix was non-urban. Urban fragment datasets could encompass any main vegetation type (typically it was forest), where the surrounding matrix was entirely comprised of intensive urban land uses. Island areas were converted into hectares. We tested for multicollinearity between predictor variables using variance inflation factors (threshold of four), and we tested for normality in each predictor. As a result *Amax* and *Sscale* were removed from subsequent analyses, while *Amin, Ascale, Smin, Ni* and *Smax* were log-transformed to induce normality.

5.4.2 Model comparison

Following Triantis *et al.* (2012) we compared 20 ISAR models (Table 5.1) using an information theoretic approach (Burnham & Anderson, 2002). The linear model was fitted using ordinary linear regression. The remaining models were fitted using non-linear regression and an expanded version of the 'mmSAR' R package model-fitting procedure (Guilhaumon *et al.*, 2010; see Triantis *et al.*, 2012, for a more detailed account). Model residuals were evaluated for normality using the Shapiro normality test, and for

homoscedasticity using Pearson correlations. The fit of a model was deemed to be satisfactory if both of these assumptions were met, and the optimisation algorithm converged; if not the fit was deemed inadequate. To avoid local minima we started the optimisation algorithm from multiple different random starting points (n=1000).

Table 5.1 The 20 ISAR models compared in the model selection. Adapted from Triantis *et al.* (2012); for further information on the various models see Dengler (2009), Tjørve (2009) and Williams *et al.* (2009). *the epm1 model can be either convex or sigmoidal in shape depending on the parameter values. The overall model rank was calculated by standardising the generality and efficiency criteria [(criterion value – mean criterion value) / standard deviation] and summing the resultant values.

No.	Model name	Model code	No. pars.	Model shape	Overall rank
1	Power	power	2	Convex	1
2	Exponential	expo	2	Convex	2
3	Kobayashi Logarithmic	koba	2	Convex	3
4	Linear	linear	2	Linear	4
5	Persistence Function 2	P2	3	Sigmoid	5
6	Monod	monod	2	Convex	6
7	Rational	ratio	3	Convex	7
8	Asymptotic	asymp	3	Convex	8
9	Extended Power 2	epm2	3	Sigmoid	9
10	Weibull-3	weibull3	3	Sigmoid	10
11	Logistic	heleg	3	Sigmoid	11
12	Morgan–Mercer–Flodin	mmf	3	Sigmoid	12
13	Beta-P	betap	4	Sigmoid	13
14	Weibull-4	weibull4	4	Sigmoid	14
15	Negative Exponential	negexpo	2	Convex	15
16	Persistence Function 1	P1	3	Convex	16
17	Gompertz	gompertz	3	Sigmoid	17
18	Power Rosenzweig	power_R	3	Convex	18
19	Chapman–Richards	chapman	3	Sigmoid	19
20	Extended Power 1	epm1	3	Convex/Sigmoid*	20

Model performance was compared using Akaike's information criterion corrected for small sample size (AIC_c; Burnham & Anderson, 2002). The smallest AIC_c value was taken to represent the single best model for a given dataset; all models <2 ΔAIC_c of the best model were considered as having similar empirical support (Burnham & Anderson, 2002). For each dataset, we calculated AIC_c weights ($wAIC_c$) to determine the probability of each model

being the best-fitting model given the set of models and concatenated them to form a model selection profile. We also recorded: whether each model provided a satisfactory fit (described above), the best fitting model, and the observed shape (linear, convex, or sigmoid) of that model fit using the sequential algorithm outlined in Triantis *et al.* (2012).

Following Triantis *et al.* (2012) we computed model generality (the proportion of datasets for which a model provided a satisfactory fit) and efficiency (the average $wAIC_c$ for all datasets in which a model provided an adequate fit). An overall model rank was then calculated by standardising each of these properties [(criterion value – mean criterion value) / standard deviation] and summing the resultant values.

For our multi-model comparative analyses we used the nonlinear implementation of the power model, but for comparison with other studies, we also fitted the logarithmic form of the power model to each dataset using standard linear regression in log–log space, recording the *c* and *z* parameters, the R^2 value, and whether the slope of the regression line was significantly different from zero. In datasets which included islands with zero species, we added one to each island richness value prior to log transformation.

5.4.3 Model performance consistency

The effect of sample size

To determine whether model ranks were consistent across the spectrum of *the number of islands (Ni)* values, model ranks were first determined for all datasets with seven or more islands (the minimum was set to seven as AIC_c could not be computed for datasets with fewer islands, see below; cf. Triantis *et al.*, 2012). This process was then repeated for all datasets with eight or more islands, and so on, iteratively up to datasets with 20 or more islands. For each model we then plotted *Ni* against model rank and fitted simple linear regressions.

Robustness

To test the sensitivity of a best-fitting model to individual data points (a criterion we termed ‘robustness’) we randomly selected 40 datasets and used a jack-knife procedure. We chose

this number because the procedure was computationally intensive and 40 represented approximately a fifth of our datasets. We hypothesised that model fits for datasets with few data points were more likely to be influenced by individual data points and therefore used a weighting system in which datasets were weighted (i.e. probability of selection) according to the number of islands they contained. For a given dataset we fitted 18 ISAR models to the complete dataset and defined the best overall model ($B_{overall}$) as that with the lowest AIC_c value. The betap and weibull4 (Table 5.1) models were not used in this analysis as the fitting process was too computationally intensive. We then removed a data point and re-ran the model selection, noting the best model for the subset (B_{trial}). If no model provided a satisfactory fit the iteration was discarded. That data point was then re-instated and a different point removed, and this process repeated iteratively until the model selection had been run for all possible sets (trials) of n-1 data points. We then calculated the percentage of these successful trials in which B_{trial} was the same as $B_{overall}$, and the percentage of trials in which a B_{trial} model shape matched the $B_{overall}$ model shape.

5.4.4 Explaining variation in ISAR patterns

Dataset characteristics

Constrained analysis of principal coordinates (CAP; Bray-Curtis dissimilarity, 9999 permutations; Anderson & Willis, 2003) was used to determine the amount of variation in the model selection profile explained in regards to (a) the best-fitting model (i.e. vectors of $wAIC_c$), and (b) the best-fitting model shape (i.e. vectors of $wAIC_c$ summed across models for each shape). We used all the aforementioned dataset characteristics, except those excluded due to multicollinearity (above), as predictor variables.

Power (log–log) model

First, we calculated summary statistics for *c* and *z*, and used boxplots and Wilcoxon rank sum tests to assess how *z* and *c* varied in response to *Tax*, *Ascale*, and *Typ*. We ran these analyses

twice: first with only parameter estimates from datasets with a significant *z* value, and second with parameter estimates from all datasets. To enable comparisons of *c* and *z* values between habitat islands (number of significant datasets =135, all datasets =207), oceanic islands (n = 125 & 193), continental-shelf islands (n = 277 & 353), and inland water-body islands (n =58 & 66) we fitted the power (log–log) model to the latter three island categories using the datasets listed in Triantis *et al.* (2012); again, running the analysis when only considering datasets with a significant *z* value, and when considering all datasets. We added 11 inland water-body island datasets found during our dataset screening process to those compiled by Triantis *et al.* (2012; see Table A3.1 in Appendix 3).

Second, we fitted two sets of linear models (LMs), using *z* and *c* as the response variables and the dataset characteristics as the predictor variables. To provide more equal sample sizes between habitat island categories we grouped mountain-top islands and urban islands into one category on the basis that they represented more isolated systems than datasets grouped under our forest island category. All additional datasets were classified as 'other'. We took the absolute value of latitude. Again, we re-ran these analyses twice: once for datasets with significant *z* values (n=133 when *z* was used as the response variable, and 133 when *c* was the response; two and three datasets were removed as outliers based on Cook's distance values, respectively), and once for all datasets. Models were compared using AIC_c. We calculated the weight of evidence (WoE) of each predictor variable by summing the Akaike weights of all the models in which a variable was included (cf. Burnham & Anderson, 2002). We used the 'dredge' function in the 'MuMIn' R package (Bartoń, 2012) to fit a complete set of models, considering all appropriate predictors. All analyses were conducted in R (version 3.0.2; R Development Core Team, 2013).

5.5 Results

From >1500 published articles, 207 habitat island datasets (Appendix 3) passed the screening procedure, comprising 121 vertebrate, 47 invertebrate, and 39 plant datasets; and 127 forest, 12 mountaintop, 16 grassland, 35 urban and 17 'other' habitat island datasets (a map of these datasets is given in Fig. A3.1 in Appendix 3). The true island datasets sourced from Triantis *et al.* (2012) included 601 datasets, comprising 193 oceanic island datasets, 353 continental-shelf island datasets and 55 inland water-body island datasets (increased to 66 with the addition of the 11 we sourced); and 233 invertebrate, 152 plant, and 227 vertebrate datasets.

5.5.1 Model comparison

AIC_c could not be computed for datasets with fewer than seven islands (cf. Triantis *et al.*, 2012), so the model comparison analyses were based on a smaller subset of datasets (Table A3.1 in Appendix 3). Of these datasets, at least one model provided a satisfactory fit in 182 datasets. The power model provided the best fit (lowest AIC_c) for 24% (N=44) of the 182 datasets. The P2 model scored best on the generality criterion, although the power model was a close second (Fig. 5.1), and the power model was the highest ranked model according to the efficiency and overall ranking criteria and was the only model to perform well by each metric (Fig. 5.1, Table 5.1). However, there is a degree of uncertainty in model performance as the mean $wAIC_c$ of the power model was only 0.17, with the second most efficient model (Kobayashi) having a mean $wAIC_c$ of 0.15 (Fig. 5.1). The more complex models performed poorly, with the weibull4, betap, and heleg models (Table 5.1) never providing the best fit (Fig. 5.1). The linear model ranked second in the number of best fits (Fig. 5.1a), but was ranked lower in generality and efficiency (Fig. 5.1bc). For the majority of datasets the observed best-fitting model shape was convex (mean $wAIC_c$ of convex models from satisfactory model fits = 0.83; linear = 0.12; sigmoid = 0.09; see Table A3.1).

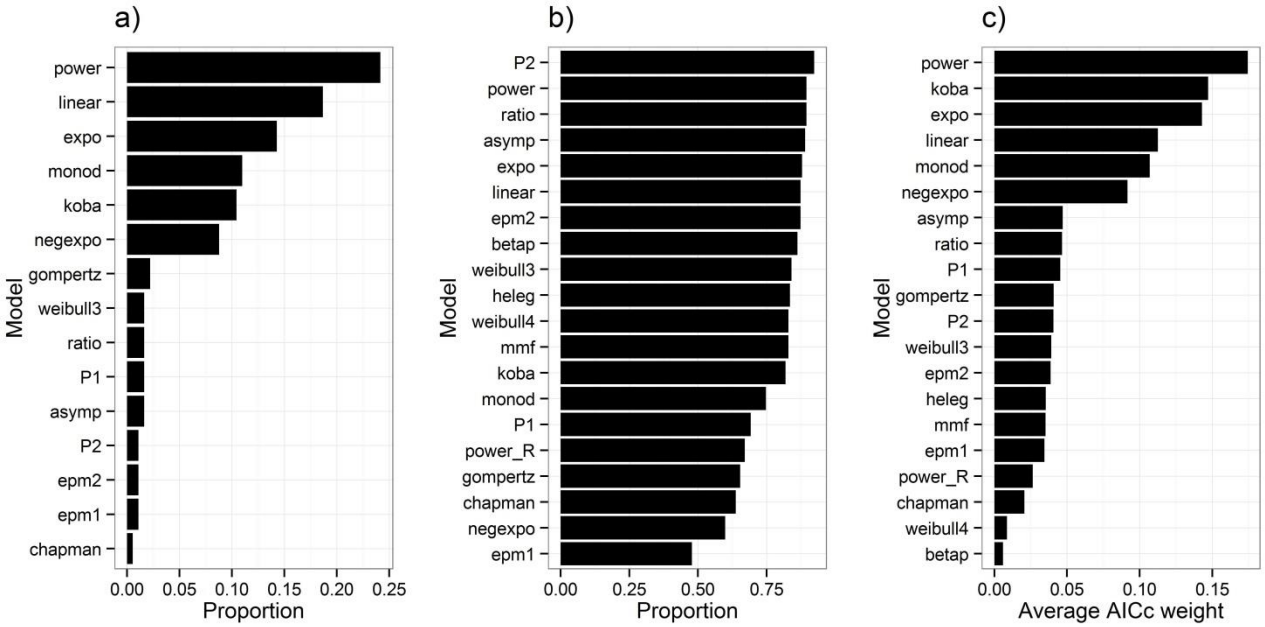


Figure 5.1 The performance of 20 species–area relationship models fitted to 182 habitat island datasets. Performance was measured in three ways: (a) the proportion of datasets for which a given model provided the best fit (i.e. had the lowest AIC_c value), (b) the proportion of datasets for which a given model provided a satisfactory fit (generality), and (c) the average AIC_c weight for datasets in which a given model provided a satisfactory fit (efficiency). The weibull4, betap, and heleg models have been omitted from (a) as they never provided the best fit to a dataset. For full model names and associated information see Table 5.1.

Sample size and robustness

The majority of model rankings were consistent across the breadth of *Ni* values. For example, the power model was ranked first for all minimum values of *Ni* (Fig. 5.2). The main exception was the linear model (Fig. 5.2), the rank of which significantly decreased with increasing minimum *Ni* (Fig. 5.2).

The model rankings appeared relatively robust to the removal of individual data points (Table A3.2 in Appendix 3). The median number of times a data point removal trial yielded the same best-fitting model as the overall best, for a given dataset, was 80%, and the median number of times a data point removal trial yielded a model with the same shape as the best overall model was 100%.

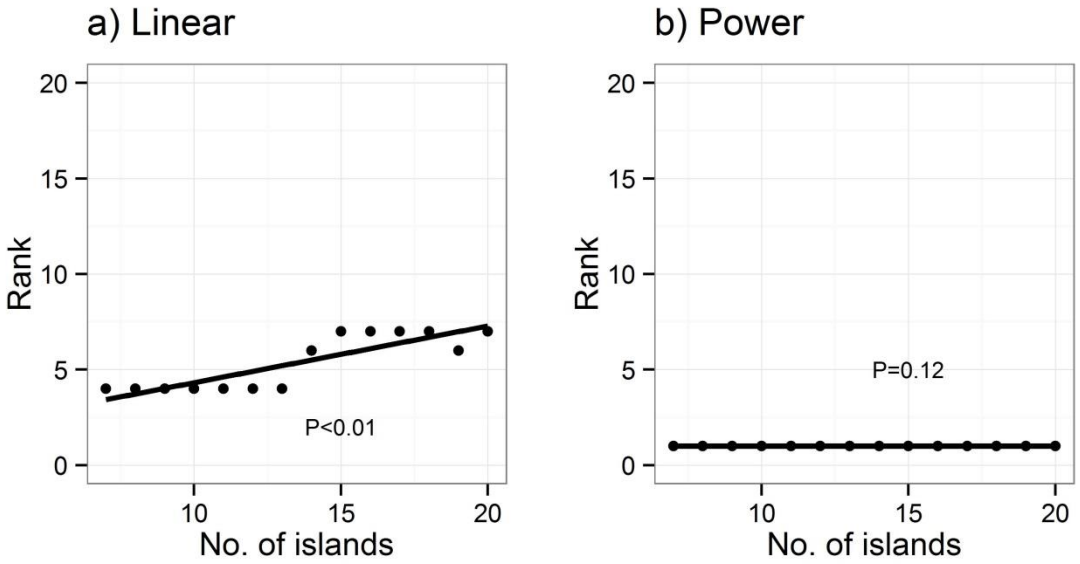


Figure 5.2 Change in model performance with the number of habitat islands in a dataset. Results for two models are displayed: (a) linear and (b) power models. The rank of a model refers to the model selection analysis in which 20 island species–area relationship models were compared using AIC_c. A model rank was determined by standardising the generality and efficiency criterion values (see 'Materials and Methods'), and adding these standardised values together. Thus, rank 1 refers to the model with the largest (larger values indicating better performance) sum of standardised generality and efficiency values. Model ranks were first determined for all datasets with four or more islands, and then for all datasets with five or more islands, and so on, iteratively up to datasets with 20 or more islands. The linear model rank results were chosen as this was the model with the biggest change in model rank with the number of islands in a dataset. A linear regression line was fitted through the points and we tested to see if the slope of the line was significantly different from zero; 'Slope sig.' refers to the *P* value of this test.

Dataset characteristics and the model selection profile

When the model selection profiles (i.e. vectors of $w\text{AIC}_c$) were used as the dependent term in the CAP analysis, the significant predictor variables (i.e. $P < 0.05$) were *Ni*, *Ascale*, *Typ*, and *Amin* (Table 5.2). However, these significant variables explained a total of only 11% of variation in the choice of best model across the 182 datasets. For model shape (i.e. vectors of $w\text{AIC}_c$ summed across models for each shape), the only significant predictors were *Ascale* and *Ni*, but they explained only 4% of the variation in best model shape across the 182 datasets.

Table 5.2 Constrained analysis of principal coordinates (CAP) results for analyses of factors explaining the amount of variation in the model selection profile, for 182 habitat island datasets. The predictor variables used in the analyses were the number of islands, island type (Typ.), the area scale (the area of the largest island in a dataset divided by the area of the smallest island), the minimum island area (Min. area), the minimum and maximum species richness values in a dataset (Min. & Max. species), taxon, and the latitude of the study area. Only predictor variables with significant effects ($P < 0.05$) are presented in each table. We used these predictors in conjunction with CAP analyses (Bray-Curtis dissimilarity, 9999 permutations) to determine the amount of variation in model selection profile explained in regards to (a) the best model, and (b) best model shape. (L) indicates predictor variables that were log transformed. The model selection results from 182 habitat island datasets were used in the CAP analyses.

a) Best model

Variable	Df	Var.	F	*P*
Number of islands (L)	1	11.81	10.1	< 0.01
Typ.	5	6.99	1.2	0.04
Area scale (L)	1	4.24	3.63	< 0.01
Min. area (L)	1	1.69	1.44	0.04
Residual	173	202.25		

b) Best model shape

Variable	Df	Var.	F	*P*
Area scale (L)	1	1.52	5.04	< 0.01
Number of islands (L)	1	0.79	2.62	0.01
Residual	179	53.93		

5.5.2 The power (log–log) model

There were 135 datasets with a significant power (log–log) *z* value (mean $R^2 = 0.62$). The median *z* value was 0.22 (Q1 & Q3=0.16 & 0.32; Fig. 5.3), and the median *c* value was 2.27 (Q1 & Q3 = 1.14 & 3.03; Fig. A3.2a in Appendix 3). In comparison, *z* values were larger for continental-shelf islands (median=0.28; Q1 & Q3= 0.19 & 0.37; Fig. 5.3a), inland water-body islands (median=0.28; Q1 & Q3= 0.19 & 0.35), and oceanic islands (median=0.35; Q1 & Q3= 0.24 & 0.49). The median *z* value for all true island categories combined was 0.29 (Q1 & Q3=0.20 & 0.40; a comparison of habitat island *z* values with those for all true island categories combined is provided in Fig. A3.3 in Appendix 3). The *z* values were significantly

lower for habitat islands than for continental-shelf islands (W= 14014, P = < 0.001), inland water-body islands (W= 3211, P = 0.048, and oceanic islands (W= 4767, P = < 0.001), according to a Wilcoxon rank-sum test. In addition, *c* values were significantly lower for oceanic islands (median=1.45; Q1 & Q3= 0.33 & 2.45; P = < 0.001; Fig. A3.2a in Appendix 3), and significantly higher for inland water-body islands (median=2.81; Q1 & Q3= 1.60 & 3.78; P = 0.02), than for habitat islands (above). Continental-shelf islands (median=2.2; Q1 & Q3= 0.80 & 3.39) had lower *c* values than did habitat islands, but this difference was not significant (P = 0.82).

Considering only datasets with a significant *z* value, and within habitat island datasets, the median *z* value was lowest for forest islands (0.20; Fig. 5.3b), and increased for urban islands (0.27), and mountain-top islands (0.30). Due to differences in sample size between categories, the only significant pairwise differences were between forest islands and urban islands (Wilcoxon rank sum test P = 0.01). The median *c* value of forest islands was significantly larger than mountaintop islands (P < 0.01; Fig. A3.2b in Appendix 3), but not urban islands (P = 0.11). The median *z* values did not significantly differ between taxa (Fig. 5.3c), while vertebrates had a significantly lower *c* value than plants (Wilcoxon tests; P = <0.01; Fig. A3.2c in Appendix 3), and invertebrates (P = 0.01).

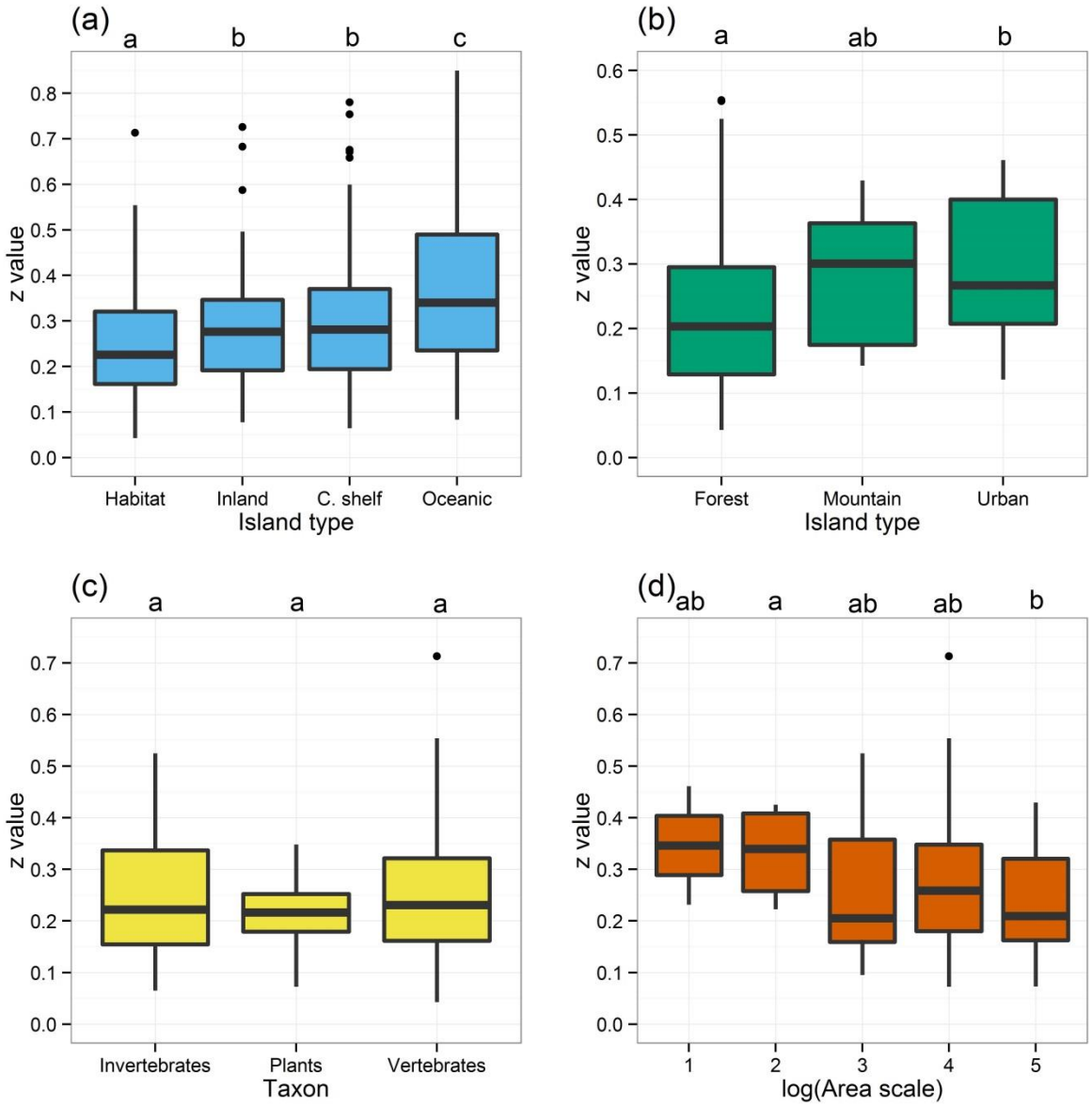


Figure 5.3 Variation in the *z* parameter of the power (log–log) model across: (a) all island types, (b) different habitat island types, (c) different major taxa, and (d) area scale (log transformed). NB (b)–(d) were calculated using only habitat island datasets. For all plots, only datasets with significant *z* values ($P < 0.05$) were included. For reasons of clarity, the box plots were constructed after omitting the small number of *z* values < 0 (all subsequent statistics were performed using the full set of data). Thus, in (a) there were 132 habitat island, 125 oceanic island, 58 inland water-body ('Inland'), and 277 continental-shelf island datasets ('C. shelf'). In (b) there were 75 forest, 12 mountaintop and 21 urban habitat island datasets. In (c) there were 26 invertebrate, 20 plant and 86 vertebrate datasets. Area scale was calculated as the area of the largest island in a dataset divided by the area of the smallest island. Area scales larger than 5 have been omitted. The box plots display the median (thick black line), the first and third quartiles (thin black box). The whiskers extend from the hinge to the highest value that is within 1.5 multiplied by the inter quartile range, of the hinge. Outliers are indicated by solid dots. Significant differences in *z* values between dataset categories are displayed as different lowercase letters above the box plots. Values that do not significantly differ between categories have the same lowercase letters.

Considering only significant *z* values, when *z* was used as the response variable in a set of LMs, the best model contained *Ascale* (Fig. 5.3d), *Ni*, *Smin* and *Smax* (Table 5.3). WoE values for these variables were high. The best model had an adjusted R^2 of 0.63. *Typ* and *Amin* were also included in some of the models <2 ΔAIC_c of the best model. Subsequent analysis of the relationship between *z* and *Ascale* revealed no clear patterns across the three taxa (Fig. A3.4 in Appendix 3). When *c* was used as the response variable, the best model (adjusted R^2 of 0.89) included *Amin*, *Ni*, *Typ* and *Smin* and each of these variables had high WoE values (Table 5.3).

When our power model analyses were re-run using parameter estimates from all datasets (i.e. including non-significant parameter estimates) our results were qualitatively similar to those based on significant parameter estimates (all results based on these data are provided in Table A3.3 and Figs. A3.5 & A3.6 in Appendix 3). For simplicity, in the Discussion, we focus solely on results based on significant parameter estimates.

5.6 Discussion

Of 20 ISAR models and for 207 habitat island datasets, the power model (nonlinear form) was the overall best model. In separate analyses using the log–log power model, we found that *z* and *c* varied between datasets, with 63% and 89% of the variation explained by various dataset characteristics, respectively.

Table 5.3 Parameter estimates for a set of the most parsimonious linear models, modelling (a) the *z* value of the power (log–log model) and (b) the value of the *c* parameter, for 135 habitat island datasets. The predictor variables included the latitude of the study site (*Lat.*; absolute value was used), area of the smallest island (*Amin*), area scale (*Ascale*; i.e. *Amax*/*Amin*), the number of islands (*Ni*.), the minimum and maximum species richness values for a dataset (*Smin* & *Smax*) the taxon (*Tax.*), and the habitat island type (*Typ.*). The best model (i.e. lowest AIC_c) and all models within ΔAIC_c of < 1.5 of the best model in (a), and within ΔAIC_c of < 2 of (b) are given in each instance. Only datasets with significant *z* values were used. The weight of evidence of each variable, calculated by summing the Akaike weights of all the models in which a variable was included, is also given. + indicates a significant effect of taxon or habitat type. A blank space indicates that a variable was not included in a model. (L) indicates predictor variables that were log transformed. The AIC_c and the delta AIC_c (ΔAIC_c) for each model selection are also presented.

a) *z* value

Model Number	Lat.	Ascale(L)	Amin(L)	Ni (L)	Smax (L)	Smin (L)	Tax	Typ.	ΔAIC_c	$wAIC_c$
1		-0.03		-0.03	0.14	-0.17			0	0.16
2		-0.04	<-0.01	-0.03	0.15	-0.17			1	0.09
3		-0.03		-0.03	0.15	-0.17		+	1.34	0.08
4		-0.04	<-0.01	-0.03	0.15	-0.17		+	1.39	0.08
Weight of evidence	0.27	1	0.45	0.84	1	1	0.38	0.38		

b) *c* value

Model Number	Lat.	Ascale(L)	Amin(L)	Ni (L)	Smax (L)	Smi n (L)	Tax	Typ.	ΔAIC_c	$wAIC_c$
1			-0.22	0.13		1.16		+	0	0.12
2			-0.22			1.14		+	0.42	0.09
3	<-0.01		-0.22	0.16		1.17		+	0.45	0.09
4			-0.22		0.08	1.09		+	1.73	0.05
Weight of evidence	0.38	0.25	1	0.57	0.28	1	0.25	0.87		

5.6.1 Model performance

It is reassuring that the power model consistently emerged as the best overall model given the preponderance of SAR studies which exclusively use this model (e.g. Watling & Donnelly, 2006). Our results are also consistent with those of Triantis *et al.* (2012) using true island datasets, in which it was found that the power model was the best-ranked model overall. The

more complex models generally performed poorly (Fig. 5.1), and as such it seems inappropriate to prefer them over simpler models for general use either in habitat island or true island studies. With regard to shape, the best-performing models were convex (mean $wAIC_c$ of 0.83). The poor performance of sigmoidal models, such as the Logistic and Gompertz models (Table 5.1) may reflect the limited range in area of most habitat island systems. Scale dependency in ISAR shape has long been debated (e.g. He & Legendre, 1996), and logistic models are theorised to be potentially appropriate only over a large range of island areas (He & Legendre, 1996; Triantis *et al.*, 2012), a prediction seemingly backed up by our results. Considering datasets in which the best fit provided an observed linear shape, the median *Ascale* value was 78, whilst the medians for datasets in which the best model provided an observed convex or sigmoidal shape were 151 and 161, respectively. In addition, *Ascale* was a significant predictor variable in the CAP analysis of the best model shape, but explained only a small amount of variance in the choice of best model shape (Table 5.2): again, possibly due to the smaller number of datasets in the larger *Ascale* categories.

In general, our predictor variables failed to explain variation in the identity and shape of the best-fitting models (Table 5.2). This may indicate that other dataset-specific factors, such as matrix properties, hunting pressure, etc., act to modulate the functional form of the ISAR (e.g. Benchimol & Peres, 2013). We have argued elsewhere that such 'confounding variables' may underpin the good fit of discontinuous piecewise ISAR models to a number of habitat island datasets (Matthews *et al.*, 2014b).

5.6.2 Interpreting the parameters of the power (log–log) model

Our study represents the largest collection of solely habitat island datasets used to date in an ISAR synthesis, and only focusing on ISAR-structured data (i.e. no nested SARs; Drakare *et al.*, 2006). Thus, it is encouraging that our reported average and range of *z* values are

consistent with values published in other syntheses that have included habitat island datasets (Table 5.4).

Table 5.4 Reported average and/or range of *z* values of isolate systems from a selection of island species–area relationship syntheses and meta-analyses. Only studies which focus on multiple datasets/island systems are included. *Represents a theorised range of *z* values, i.e. not results from a synthetic analysis. +This number includes datasets that report *z* values derived from nested SAR data, rather than ISAR data; the average *z* value reported in the table only relates to the ISAR structured datasets.

Dataset	Island type	No. of datasets	Taxon studied	Reported average and/or range of *z* values
This study	Habitat islands	207	Multiple taxa	Median = 0.22 (Q1 & Q3=0.16 & 0.32).
Connor & McCoy (1979)	Multiple island types	90	Multiple taxa	Mean = 0.31; standard deviation = 0.23 (range= -0.28 to 1.13).
Drakare *et al.* (2006)	Multiple island types	794+	Multiple taxa	Average = 0.24 (range = 0 to *c.* 1)
Rosenzweig (1995)	True islands	*	*	Range: 0.25-0.33 *
Sala *et al.* (2005)	Oceanic and mountaintop islands	26	Vascular plants	Mean = 0.34; standard deviation = 0.14
Sólymos & Lele (2005)	Multiple island types	94	Multiple taxa	Mean = 0.23 (90% confidence limits of 0.06 & 0.41).
Triantis *et al.* (2012)	True islands	601	Multiple taxa	Mean = 0.32; standard deviation = 0.16 (range = 0.06 to 1.31).
Watling & Donnelly (2006)	Habitat islands & true islands	118	Multiple taxa	Habitat islands: mean = 0.20. True islands: mean = 0.26.

A central aim of much SAR research has been to determine whether the *z* parameter is biologically interpretable (Connor & McCoy, 1979; Rosenzweig, 1995; Triantis *et al*., 2012). One particular area of interest has focused on the idea that more permeable matrices surrounding islands will result in lower *z* values as individuals of certain taxa can more easily disperse between islands, leading for example to rescue effect processes (Watling &

Donnelly, 2006; Whittaker & Fernández-Palacios, 2007). Hence, it has been argued that habitat island systems should, in general, have lower *z* values than true island systems (see Table 5.4; MacArthur & Wilson, 1967; Rosenzweig, 1995). In illustration, Mendenhall *et al.* (2014a) show that forest fragments in an agricultural matrix in Costa Rica have lower extinction rates and shallower ISAR slopes than nearby true islands in Panama. Our results indicated that in general habitat islands (median $z = 0.22$) do indeed have lower *z* values than both continental-shelf islands (0.28), and oceanic islands (0.35; Fig. 5.3a), most likely reflecting the dominant processes involved in island biota assembly/disassembly (Triantis *et al.*, 2012). We also found that median *z* increased from forested islands (0.20) to mountain-top islands (0.30) and urban islands (0.27; Fig. 5.3b), indicating that *z* seemingly responds to matrix type (see also Watling & Donnelly, 2006).

ISAR *z* values were also found to be affected by *Smin*, *Smax*, *Ascale* and *Ni*. For *Ni* the effect was negative, meaning *z* decreased with increasing *Ni*. This finding is consistent with our model selection results as we found that the linear model performed better as the number of islands in a dataset decreased. Thus, it seems that in datasets with low *Ni* values there is more of a linear shape to the ISAR in arithmetic space, which in turn often results in a steeper ISAR in log–log space. That *z* decreased with increasing *Ascale* seems to imply that the most dramatic increase in richness occurs over a low range of island areas and that beyond a certain size of island the gain in richness from additional area becomes slight. The negative effect of *Smin* on the *z* value, coupled with the positive effect of *Smax*, makes sense as together these two variables represent the range in species richness within a dataset. If we assume that *Smin* typically occurs on a small island, while *Smax* occurs on one of the largest islands, then decreasing *Smin* and increasing *Smax* will necessarily result in a steeper slope in log–log space.

The lack of difference in *z* values between taxa is interesting as studies that focused primarily on true islands have found significant inter-taxa differences (Sólymos & Lele, 2012; Triantis *et al.*, 2012), whilst a separate analysis of habitat island datasets also found no significant difference (Watling & Donnelly, 2006). We are unsure as to the precise causes of this result, but it is possible that our taxonomic division (i.e. plants, vertebrates and invertebrates) is simply too coarse, with too much internal variation in dispersal powers within the groupings (Aranda *et al.*, 2013); while dividing species into smaller taxonomic groups leads to sample sizes too small to allow any reasonable inference. Interestingly, vertebrates had significantly lower *c* values than plants and invertebrates. The *c* parameter is often overlooked and has been argued to reflect a number of different properties; for example, the average size of the most common species, a measure of carrying capacity, and a scale-independent measure of diversity (Gould, 1979; Whittaker & Fernández-Palacios, 2007; Triantis *et al.*, 2012). Our finding that vertebrates have a significantly lower *c* value supports the latter two hypotheses as vertebrates, in general, are less diverse than plants and invertebrates.

5.6.3 Conservation implications and conclusions

Habitat loss is the biggest driver of the current wave of species extinctions (Schipper *et al.*, 2008). It is thus essential that conservation biogeography develops an accurate and coherent methodology for using the ISAR to predict extinctions resulting from habitat reduction (Sala *et al.*, 2005). Deriving any generalities regarding the form of the ISAR in habitat islands, alongside deriving appropriate parameter values to use in prediction exercises, are essential components of this research programme (Halley *et al.*, 2013). The functional form of the model is a particularly important consideration as the different forms of the ISAR curve translate to varying forms of their respective first derivatives, and thus different species

extinction rates for a given amount of habitat area (Fig. A3.7 in Appendix 3 illustrates this issue using a convex and sigmoidal model).

The large number of datasets used herein allows us to provide at least partial answers to a number of the questions that have been posed in connection with applied use of the SAR (e.g. Connor & McCoy, 1979; Rosenzweig, 1995; Laurance, 2008). For example, in a seminal critique of the use of the SAR in biogeography, Connor & McCoy (1979), list, amongst others, (1) the unreserved use of the power model, and (2) the use of canonical *z* values of 0.25 derived from traditional island theory (e.g. Preston, 1962), as widespread incorrect assumptions. Regarding issue (1), we have shown that the power model is markedly the best overall model according to our criteria and thus is in general a sensible choice (but see Smith, 2010). Regarding issue (2), we found that *z* varies according to a number of different properties, such as island type, a variable which is linked to matrix type: as the intervening matrix becomes more hostile/less permeable, the *z* value increases. This is particularly concerning as the majority of past studies have used *z* values from a set range (0.18–0.25) to predict extinctions (see Halley *et al.*, 2013). Our results indicate such values may be appropriate for only a subset of systems; while the median *z* (log–log model) value when all datasets with a significant model fit (n=135) were considered was 0.22, 51 datasets had *z* values >0.25. Perhaps a better approach when the aim is to predict extinctions would be to adjust *z* values according to a particular context/set of site specific conditions (as Triantis *et al.*, 2010). This need not introduce large amounts of complexity, and even very basic adjustments have had positive results (Pereira & Daily, 2006; Koh *et al.*, 2010; Mendenhall *et al.*, 2014a, b). There is a need for more research in this area, and further broad-scale tests of previously published 'calibrated' and 'countryside biogeographic' SAR models.

It is important to recognise that determining the functional form of the ISAR in habitat island systems and selecting the most appropriate model, or set of models, are but

steps towards improving the accuracy of ISAR extinction predictions (Matias *et al.*, 2014). A number of other factors are also relevant. For example, (i) variables other than area have been found to modulate the SAR (e.g. hunting pressure; Benchimol & Peres, 2013); (ii) the ISAR is often unable to accurately predict the total richness of a set of isolates; (iii) and there may be interactions between the effects of decreasing patch area and increasing patch isolation (Hanski *et al.*, 2013). A further issue is that the majority of studies, including the present analysis, include all sampled species in SAR calculations. However, in a recent paper we have shown that habitat specialists and generalists respond discordantly to habitat insularization: the first derivative of a multi-model ISAR curve was generally steeper for specialists than for generalists (Matthews *et al.*, 2014a; see also Banks-Leite *et al.*, 2012). Thus, the inclusion of generalist species in SAR calculations may act to mask the impact of habitat loss on specialists.

It is necessary to highlight two potential issues with our analyses. First, whilst island type in this study is likely coupled with matrix type, it would have been preferable to include more precise information on matrix type in our models. The categorization used was necessarily coarse given the lack of detailed information provided in many source papers, but the fact significant differences were found suggests this may be a productive avenue for further work in the future. Second, as with any meta-analysis the inclusion of multiple datasets from different sources mixes studies that had varying aims, sampling protocols and sampling effort, which may affect the inferences drawn (Whittaker, 2010). It is thus possible that variations in data quality have influenced our results. To counter this issue we read and checked the methods of each source papers before accepting a dataset. Having done so we saw no reason to assume that variation in sampling effort would be biased towards any particular island type or taxon. However, future work might usefully be undertaken, for example, to determine whether studies have consistently under-sampled the larger islands in a

dataset (which might affect conclusions on ISAR curve shape, for example). See Appendix 3 for a more detailed discussion of data quality issues in SAR studies.

Due to constraints on resources and data availability, the analysis of SARs remains one of the most important analytical methods in the conservation biogeographer's tool kit. Thus, it is essential that studies attempt to synthesise information across systems in order to produce general guidelines. Based on the results of this study, we suggest the following guidance for using the ISAR with habitat island data: (a) if only one model is to be used, the power model is a sensible choice, (b) convex models should generally be preferred to linear and sigmoidal models, particularly if very large islands are not the focus of study, (c) depending on the aim of the study, calibrating the *z* value of the power model based on system specific characteristics will improve predictive accuracy, and (d) as the software is freely available (Guilhaumon *et al.*, 2010), it can be enlightening to fit a selection of models and use a multi-model inference approach, especially for use in conservation applications.

CHAPTER 6

Island species–area relationships and species accumulation curves are not equivalent: an analysis of habitat island datasets

Paper in review in Global Ecology & Biogeography

6.1 Preface

This chapter is the third empirical chapter, and examines the differences in the slope of island species–relationships (ISARs) and species accumulation curves (SACs) using a large set of habitat island datasets. For each dataset, the ISAR and a single form of SAC were constructed using the habitat island data. The power (non-linear) model was fit to both SAR types, and the difference in the *z* values of the power model taken to represent the difference in slope between the two SARs. The ability of the ISAR to predict the richness of larger areas was also tested. It was found that SAC was generally steeper than the ISAR, an observation that was linked to the low degree of nestedness in most archipelagos.

Authors: Thomas J. Matthews, Kostas A. Triantis, François Rigal, Michael K. Borregaard and Robert J. Whittaker

The co-author contributions were as follows. T.J.M., K.A.T. and R.J.W. conceived the ideas; T.J.M. collected the data; T.J.M. ran the analyses; and T.J.M. and R.J.W. wrote the paper with a significant contribution from K.A.T, M.K.B and F.R. All authors discussed the results and commented on the manuscript.

6.2 Abstract

Aim Despite the existence of a number of published species–area relationship (SAR) typologies, numerous SAR studies fail to distinguish adequately between the island species–area relationship (ISAR) and the species accumulation curve (SAC). We undertake a meta-analysis of 97 habitat island datasets to determine how the slope of the SAR varies between the ISAR and SAC, for the same dataset. We also assess how well the ISAR can predict the total number of species in an archipelago.

Locations Global

Methods For each dataset we constructed the ISAR and the SAC from the same island data. We then compared differences in the rate of change in richness with area between the ISAR and the SAC using the difference in the *z* value (*z*Dif) of the power (non-linear) SAR model. We used boosted regression tree and simulation analyses to investigate the effect of nestedness and other variables in driving observed differences in slope between ISARs and SACs.

Results Comparison of the model fits revealed substantial differences in the rate of change in species richness with area between the ISAR and the SAC, for the same dataset. The SAC was steeper than the ISAR in 77% of datasets according to *z*Dif. Across multiple datasets, the difference in slope was primarily driven by the degree of nestedness in the system, although other variables (e.g. the number islands in a dataset) were also important. We also found that the ISAR was often a poor predictor of the number of species in an archipelago.

Main conclusions The slope of the SAR for the same dataset varies systematically between the ISAR and the SAC. These results highlight a number of issues with the assumptions that certain SAR studies make. For example, the ISAR was a poor predictor of archipelagic richness in datasets with a low degree of nestedness, and thus caution should be employed when using the ISAR for extrapolation and prediction purposes in habitat island systems.

6.3 Introduction

While the 'species–area relationship' (hereafter SAR) is generally used to refer to any relationship between the number of species in a unit being sampled and the area of the unit, a number of distinct forms of SAR have been recognised (Scheiner, 2003; Gray *et al.,* 2004a; Whittaker & Fernández-Palacios, 2007). As a first step these different forms can be dichotomised into island species–area relationships ('ISARs') and species accumulation curves ('SACs') (Whittaker & Matthews, 2014; but see Scheiner 2003). The ISAR is a term used to refer to Scheiner's (2003) type IV curve, whereby the number of species occurring within each of a set of islands is analysed as a function of the area of each island (Triantis *et al.,* 2012). The SAC refers to plots of increasing cumulative species number with increasing sampling effort/area (Whittaker & Matthews, 2014). Published typologies recognise multiple types of SACs denoting the various ways in which sampling effort can be deployed (see Scheiner's, 2003, type I-III curves). Furthermore, the SAC can be constructed in multiple ways that vary the order in which samples are accumulated to build the curve: in order of increasing or decreasing area, or of increasing or decreasing species richness, etc. (e.g. Quinn & Harrison, 1988; Fischer & Lindenmayer, 2002; Watson *et al.,* 2009; Gavish *et al.,* 2012; and see our Appendix 4). In this study, rather than focus on a particular SAC ordering rule, we use a randomly constructed SAC (Ugland *et al.,* 2003; Gray *et al.,* 2004a, b), which is an average of multiple SAC permutations.

Despite the availability of published SAR typologies (e.g. Rosenzweig, 1995; Scheiner, 2003; Gray *et al.,* 2004a; Dengler, 2009), incorrect assumptions are quite often made concerning SARs. These can be condensed into three broad issues. First, some studies implicitly assume ISARs and SACs to be equivalent curve types. As noted by Hui (2008, p.327), there is a "general confusion about species–area curves and species accumulation curves", which can result in authors using the incorrect SAR type for a given purpose. For

example, Chong & Stohlgren (2007) analysed the contribution of 17 different vegetation types to regional diversity in the Rocky Mountain National Park. The authors constructed models akin to ISARs using subplot data (i.e. they only considered counts of species in each subplot rather than any measure of species overlap) in order to predict the total richness in the different vegetation types. However, the predictions are generally inaccurate because Chong & Stohlgren (2007) should have used SACs in their calculations (see Hui, 2008). A further example can be found in the SAR typology of Rosenzweig (1995). Rosenzweig argued that there are three primary scales of SARs (islands, intraprovince, interprovince). Thus, his typology was primarily interested in the effect of scale. However, as highlighted by Whittaker & Fernández-Palacios (2007), the typology incorporates both ISAR and SAC formulations, essentially assuming their parameters are inter-comparable.

Second, the ISAR is often used in conservation applications for forward and backward extrapolation (Stohlgren *et al.*, 1997; van Vuuren *et al.*, 2006; Chong & Stohlgren, 2007), even in cases where the data are available to construct SACs. Theoretically the SAC should be more useful in extrapolation as the ISAR is silent on species overlap (Quinn & Harrison, 1988; Crist & Veech, 2006). It has been suggested that generally the end point of the ISAR is less than the total number of species in the system (e.g. Crist & Veech, 2006). However, few studies have examined the validity of extrapolations of ISARs using multiple datasets. One exception is the study of Santos *et al.* (2010), who used island datasets to test how closely the cumulative species total of the archipelago ('archipelagic point') was predicted by the ISAR of the archipelago's constituent islands. While the archipelagic point often fell close to the ISAR of the constituent islands, the study was focused predominantly on oceanic and continental shelf islands and thus its relevance to habitat island systems (for which extrapolation is more common) is unknown.

Third, it is sometimes assumed that the slope of the ISAR represents the degree of species overlap between islands (see discussion in Connor & McCoy, 1979; Rosenzweig, 1995; see also studies listed in Wilson & Shmida, 1984; Veech *et al.*, 2002; Crist & Veech, 2006). Such a statement only necessarily relates to the SAC (Gray *et al.*, 2004a; Crist & Veech, 2006), although again the relationship between the slope of the ISAR and SAC, and a metric of species overlap and/or nestedness has not been tested using multiple habitat island datasets.

Several of the aforementioned issues have previously been identified. The debate in this journal between Scheiner (2003, 2004) and Gray *et al.* (2004a, b) over the differences between ISARs and SACs is a case in point. However, the issues have generally been approached theoretically, or by using a handful of empirical datasets, and hence a number of authors have argued that to fully understand the implications of the differing structures and form of ISARs and SACs there is a need for comparative analyses of different SAR structures (e.g. Drakare *et al.*, 2006; Whittaker & Fernández-Palacios, 2007; Mazel *et al.*, 2014).

In addition to exploring how ISAR and SAC curves differ, it is also important to identify the mechanisms driving any observed differences. Intuitively, the slope of the SAC should be related to the degree of nestedness (i.e. species nestedness between isolates) observed in the system (Gray *et al.*, 2004a, b; Watson *et al.*, 2009), whereas there appears to be no inherent reason for the slope of the ISAR to vary with the degree of nestedness. Studies have shown mathematically how the slope of the SAC is a function of the degree of species overlap between areas (e.g. Tjørve & Tjørve, 2008), but there has been no analysis of exactly how the degree of overlap and/or nestedness in an archipelago constrains the difference in slope between the SAC and ISAR, e.g. at what level of nestedness does the ISAR become steeper than the SAC? Furthermore, there may be additional mechanisms involved. For example, the nature of the matrix surrounding habitat islands could affect the difference

between SACs and ISARs due to its effect on species overlap between islands: a more permeable matrix is predicted to result in a greater degree of species overlap between islands as inter-island dispersal is increased (Whittaker & Fernández-Palacios, 2007; Matthews *et al.*, *in press*).

Constructing SACs from island data has previously been attempted (e.g. Quinn & Harrison, 1988; Rosenzweig, 1995; Fischer & Lindenmayer, 2002) but rarely have ISARs and SACs been compared using such an approach. The primary example is that of Rosenzweig (1995, pp.19–20), who, using a single archipelago as evidence, argued that the slopes of a SAC and of an ISAR plotted using the same data from a set of islands will not differ much. Furthermore, we are unaware of any analysis of the differences between the two main SAR forms incorporating a large number of datasets (but see Crist & Veech, 2006). In this study we undertake a synthetic analysis of 97 habitat island datasets to determine how the slope of the SAR varies between the ISAR and the SAC, for the same dataset. We test the following hypotheses: (1) that the slope of the ISAR will differ from the slope of the SAC, for a given dataset; (2) the magnitude of this difference will primarily be related to the degree of nestedness in the system, such that in low nested systems the SAC will be steeper than the ISAR and as the degree of nestedness increases this situation will gradually reverse. In addition we also examine how well other mechanisms might affect the examined differences between the ISAR and SAC, and the ability of the ISAR to predict the total richness of the archipelago (i.e. a hypothetical island that has an area equivalent to the summed area of all the islands; cf. Santos *et al.*, 2010).

6.4 Methods

6.4.1 Data collection

We searched three abstracting services for relevant habitat island studies and datasets between May 2010 and August 2013: JSTOR (1913–2003); ISI Web of Knowledge (1980–

2013); and BIOSIS Biological Abstracts (1980–2003). We used search keywords, encompassing different combinations of 'habitat islands', 'species richness', and 'forest fragments'. We searched through the reference lists of papers found in these abstracting services for additional studies. Certain datasets were obtained from the authors of the source papers, whilst others were supplemented with additional data from the source paper authors. Following Matthews *et al.* (2014b), our criteria for inclusion of source data were as follows: (i) habitat islands were defined as discrete patches of habitat surrounded by contrasting matrix habitat (we also included a small number of datasets consisting of protected areas in which the contrast between the islands and the intervening matrix was not so pronounced); (ii) the number of habitat islands within each dataset was at least five; and (iii) the area of each habitat island and was listed, and the species lists provided. For the purposes of this study, we considered a few island systems within an aquatic matrix (e.g. forest fragments within a reservoir) as habitat islands, as the range of island areas and dominant assembly processes are more similar to habitat islands than oceanic islands. For each dataset we recorded: the geographic location, the taxon studied, the habitat island type (forested or non-forested), the latitude and longitude of the study extent (for some datasets this was an estimate as the data were not presented in the source papers), the range of species richness and island sizes, and a classification of the predominant matrix type in each study (see Appendix 4 for more detailed information).

6.4.2 Data formatting and SAR construction

For each dataset we constructed two types of SAR. First, we created the standard ISAR using the raw island area and richness values. Second, we also constructed a smoothed random SAC (herein, "SAC") using a simple bootstrap procedure. The data points were generated by

randomly selecting the order of island addition into the SAC at each iteration, and taking the average data point values of 5000 iterations. For example, for the second data point of the SAC of a dataset with six islands, the area and species richness of the data point were calculated as the mean of the addition of 5000 random combinations of two islands, out of the six (see Appendix 4 for further details).

Our aim was to compare the slope of the ISAR with the slope of the SAC, for the same dataset. The simplest way to do this is to compare the *z* values of the power model fitted to both sets of data. However, this method is arguably only appropriate if the power model provides a reasonably good fit to both sets of data. Thus, we devised a set of criteria to select suitable datasets for analysis. First, we fitted the power (non-linear) SAR model to the two different SAR types for each dataset using non-linear regression and the 'mmSAR' R package (Guilhaumon *et al.*, 2010), and recorded the two parameters (*c*, *z*), the R^2, and whether the slope was significantly different from zero. A dataset was deemed satisfactory according to this process if both *z* values (i.e. the *z* values generated from fitting the power model to the SAC and ISAR data) were significant, and if the R^2 was equal or greater than 0.5 in both cases. Second, as the observed shape of the power (non-linear) model is generally convex, it was necessary to determine whether a model with a different shape provided a better fit to our data. Thus, we fitted a set of 20 ISAR models to the ISAR and SAC data from each dataset using a modified version of the fitting algorithm within the mmSAR R package (see Triantis *et al.*, 2012). We considered the model with the lowest Akaike's information criterion corrected for small sample size (AIC_c), as providing the best fit (Burnham & Anderson, 2002). AIC_c could not be calculated for datasets with fewer than seven islands and thus these datasets were classified as unsatisfactory. Models which produced inadequate fits (e.g. due to non-normal residuals; cf. Triantis *et al.*, 2012) were discarded. We then determined the observed shape of the best model fit (convex, sigmoid or linear) according to

the algorithm outlined in Triantis *et al.* (2012; see Appendix 4 for an expanded summary). For each dataset, if the observed shape of either of the best model fits (i.e. to the ISAR and SAC data) was linear or sigmoid, the dataset was discarded. In sum, to be classified as satisfactory, both the ISAR and SAC forms of each dataset needed to have: (a) significant *z* values from the power model fits, (b) R^2 values >0.5 from the power model fits, and (c) observed convex fits, when considering the best model from a set of twenty competing models.

For these satisfactory datasets, the difference between the SAC *z* value and the ISAR *z* value (zDif) was used as our response variable. We also re-ran our analyses using the zDif values from all datasets (i.e. those deemed satisfactory and unsatisfactory).

6.4.3 Relative influence of the predictor variables

To determine factors potentially accounting for between-dataset differences in zDif, we fitted a boosted regression trees model (BRT) using the functions provided inElith *et al.* (2008). In contrast to traditional regression, which results in a single best model, BRT combines a large number of tree models in order to increase predictive performance (Elith *et al.*, 2008). Recent work has shown that using weight of evidence values (Burnham & Anderson, 2002) within a linear modelling framework to determine the relative importance of predictor variables is flawed (Galipaud *et al.*, 2014). BRT offers a useful intuitive and robust alternative (for further discussion, see: Elith *et al.*, 2008).

In the BRT analysis, zDif was used as the response variable and the dataset characteristics listed above were used as predictor variables. As a further predictor variable, we also calculated the degree of nestedness in each dataset using the NODF metric (Almeida-Neto *et al.*, 2008) and the 'vegan' R package (Oksanen *et al.*, 2013). NODF was calculated using the maximally packed matrix (i.e. row and column total were allowed to vary), and any sites containing zero species were removed prior to calculation. Cross-validation

functionality was used to determine the optimum number of trees. We experimented with a variety of tree complexity, learning rate and bag fraction parameter values in order to minimise the minimum predictive errors, selecting final values of 5, 0.001 (0.01 when using all datasets) and 0.5 respectively. These values resulted in an optimal number of 7050 fixed trees using only the satisfactory datasets, and 4500 trees when all datasets were used. The error structure was considered to be Gaussian. No outliers were removed when running the BRT analysis using the satisfactory datasets, but two outliers had to be removed when using all datasets to induce normality in *z*Dif. The predictive power of the model was assessed through cross-validation, whereby the fitted model was tested on withheld portions of the data. In this cross-validation procedure, the data were randomly divided into 10 even subsets. One subset was then withheld and the model fitted to the data in the remaining 9 subsets, and the predictive accuracy assessed using the withheld data. The predicted response values were then correlated with the observed response values. This process was then repeated across all ten subsets, and a mean correlation coefficient taken.

The relative influence of the predictor variables was again calculated using the functions in Elith *et al.* (2008), which weight the number of times a predictor is chosen for splitting by the squared improvement to the model due to each split. This weighted value is then averaged across all tress and the relative influence of each predictor is then scaled such that the sum of all values equals 100. Larger values mean a predictor has a larger influence on the response. Partial dependence plots were also used to assess the effect of a predictor after accounting for the effects of all other model predictors (Elith *et al.*, 2008). In partial dependence plots the y-axis shows the effect of a predictor variable on the response variable after accounting for the effects of the other model predictors (i.e. the marginal effect of the predictor).

6.4.4 Simulation analyses

Our BRT analyses revealed NODF to be the most important variable (see Results). Thus, to further examine any effect of nestedness on the difference between the slope of the ISAR and of the SAC, we adopted a simulation approach. First, we simulated a set of six islands with area and species richness values conforming roughly to a convex ISAR (area= 1, 5, 10, 15, 20, and 30; containing 10, 12, 14, 16, 17 & 18 species respectively; herein 'Sim1'). We fitted the power model (non-linear) to the ISAR-structured form of Sim1 and recorded the model parameters. We then simulated a perfectly nested (i.e. NODF=100) presence/absence matrix using the data characteristics of Sim1 (i.e. island number and species richness) and used this matrix to construct the SL (small-large) SAC (see Appendix 4); fitting the power model to the resulting curve. We used the SL SAC in this analysis, as constructing the Ran SAC for this many matrices would have been too computationally intensive. We then altered the presence/absence matrix (without changing the area and richness values) to change the degree of matrix nestedness, by randomly shuffling the presences along the matrix rows (i.e. sites) using the commsimulator function in the vegan R package and discarding any matrix which had already been simulated. The total number of species in the species pool was set to 40 (i.e. the total possible number of species included in a presence-absence matrix), and we allowed gamma diversity to change between matrices, but the alpha diversity of each island remained constant. This permitted us to create matrices with substantially different levels of nestedness. However, as this meant that the number of columns (total number of species) varied between matrices we used the NODF by rows value as our measure of nestedness. For each accepted matrix, we constructed the SAC curve and fitted the power model. We started the random shuffling from different starting points (i.e. different initial presence-absence matrices), used a variety of community simulation algorithms (i.e. R0, R1 and R2) and repeated this process

iteratively for 12,000 runs, to cover a wide range of NODF values. We then examined *z*Dif values for each iteration, plotting *z*Dif as a function of nestedness.

6.4.5 Determining the degree of deviation of the archipelagic point

To determine whether the archipelagic point of a dataset deviated from the ISAR of the constituent islands, we followed the method and nomenclature outlined by Santos *et al.* (2010). We also use 'archipelagic point' to refer to the total area and richness of the corresponding set of habitat islands (i.e. the archipelago). For each dataset, we fit the power (log–log; base 10) SAR model to the ISAR structure and derived the fitted values of the model for each island ('SIpred)'. The log–log version of the power model was used in this analysis as it was the model used by Santos *et al.* (2010) and thus allowed us to compare our results with theirs. Also following Santos *et al.* (2010), we added 0.01 to each species richness value to avoid zero values. We then calculated the archipelagic point as the cumulative total area of all habitat islands in the dataset plotted against the cumulative species richness total ('SAobs'); and predicted the number of species in the archipelagic point ('SApred') using the ISAR model of the constituent islands. We followed Santos *et al.* (2010) and calculated the archipelagic residual ('ArcRes') as the absolute difference between log(SAobs) and SApred, standardised by log(SAobs) (see Appendix 4 for details). We noted whether the ISAR over- or under- predicted richness in the archipelagic point for each dataset. This ArcRes methodology represents a simple metric with which to describe how well the ISAR predicts the archipelagic richness.

To determine if any dataset characteristics (above) could explain variation in ArcRes between datasets we repeated our BRT analyses (learning rate of 0.01) using ArcRes as the response variable. We only used ArcRes values from datasets in which the power (log–log) provided a significant fit (cf. Santos *et al.*, 2010). Nine outlying points required removing to induce normality, as there was a long left-hand tail in the distribution of ArcRes values. The

optimal number of fixed trees was 2200. All analyses and simulations were conducted in R (version 3.1.1.; R Development Core Team, 2014).

6.5 Results

We screened over 1000 published articles, of which 97 were deemed suitable for analysis (Table S2 in Appendix 4), comprising 69 vertebrate, 20 invertebrate, and 8 plant datasets. According to our three additional dataset acceptance criteria, 50 datasets were deemed satisfactory for the *z*Dif analyses.

6.5.1 Differences in the slope of ISARs and SACs

Considering only the satisfactory datasets (n = 50), the *z* of the SAC was greater than the *z* of the ISAR for 32 (Fig. A4.2 in Appendix 4 and example in Fig. 6.1), while the opposite was true for 18 cases. As to be expected, the power model explained a larger amount of variance for the SAC (mean R^2 = 0.97) than for the ISAR (mean R^2 = 0.76). Considering all datasets, the *z* value was larger for the SAC for 75 of the 97 datasets.

6.5.2 Boosted regression tree results

When *z*Dif values from satisfactory datasets were used as the response variable, NODF was the most important explanatory variable (relative influence value of 24.6; see Table 6.1). Both the minimum number of species and the number of islands also had relatively high influence values (Table 6.1, Fig. 6.2a,b). In terms of the overall predictive performance of the model, the mean coefficient of the correlation between the observed and predicted response values was 0.63 when based on the cross-validation data. Results were qualitatively similar when considering *z*Dif values from all datasets (mean correlation coefficient = 0.68).

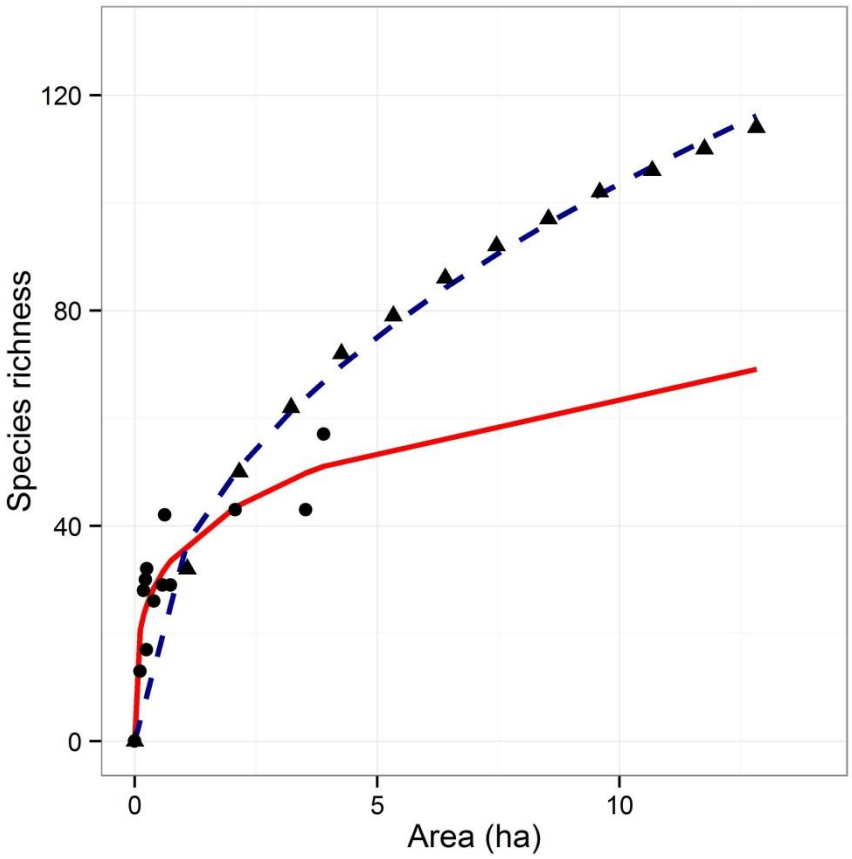


Figure 6.1 Power model fits constructed from the same habitat island dataset: Gavish *et al.* (2012; invertebrates in a study system from Israel; number of islands = 12; number of species = 114). The fits of the power model (non-linear) to the ISAR data (solid circles; solid line = fit of the model) and the Ran SAC data (solid triangles; dashed line) are shown. For the ISAR model fit, the power model has been extrapolated to the total cumulative area of all habitat islands in the system using the parameters derived from model fitting process. Both curves have been forced through the origin. The *z* values of the models in (b) are 0.25 for the ISAR model, and 0.46 for the SAC model.

As predicted, further analysis revealed that the less nested the system, the greater the *z*Dif values (see Fig. 6.2a and Fig. 6.3a). In addition, the *z* value of the SAC was significantly related to the NODF value (Fig. 6.3b), whereas there was no relationship between the *z* of the ISAR and NODF (Fig. 6.3c).

Table 6.1 The relative contributions (%) of predictor variables for boosted regression tree models developed using cross-validation. The model was fitted using two different response variables: *z*Dif (number of datasets = 50; 7050 fixed trees), and ArcRes (number of datasets = 66; 2300 fixed trees). *z*Dif is the difference between the *z* value of the SAC and the *z* value of the ISAR, for each dataset. ArcRes is the standardised difference between the cumulative species richness total of the archipelago and the predicted the number of species in the archipelagic point using the ISAR model of the constituent islands. A set of dataset characteristics were used as the predictor variables (see Materials and Methods): Min = minimum, Max = maximum, No = number of, Ar. = area, Isl. = island, Sp. = species. The nestedness of the archipelago was measured using the NODF metric (maximally packed matrix). Island type was a categorical variable indicating whether a dataset was a forested or a non-forest island, and taxon was a categorical variable indicating whether a dataset was a plant, vertebrate or invertebrate dataset. Matrix was a variable indicating the permeability of the surrounding matrix type. Variables are ordered by their relative influence (Rel. influence) in each case.

*z*Dif		ArcRes	
Variable	Rel. Influence	Variable	Rel. Influence
NODF	24.61	NODF	30.11
Min. Sp.	17.94	No. Sp.	16.62
No. Isl.	17.12	Min. Sp.	13.60
Ar. Ratio	13.15	Ar. Ratio	8.00
Max Ar.	9.80	Latitude	7.15
Longitude	4.53	Longitude	6.33
Latitude	4.52	Max Ar.	5.31
No. Sp.	2.78	Max Sp.	3.44
Max Sp.	2.14	No. Isl.	3.40
Matrix	1.43	Taxon	2.20
Min. Ar.	0.99	Min. Ar.	2.06
Isl. Type	0.88	Isl. Type	0.92
Taxon	0.11	Matrix	0.86

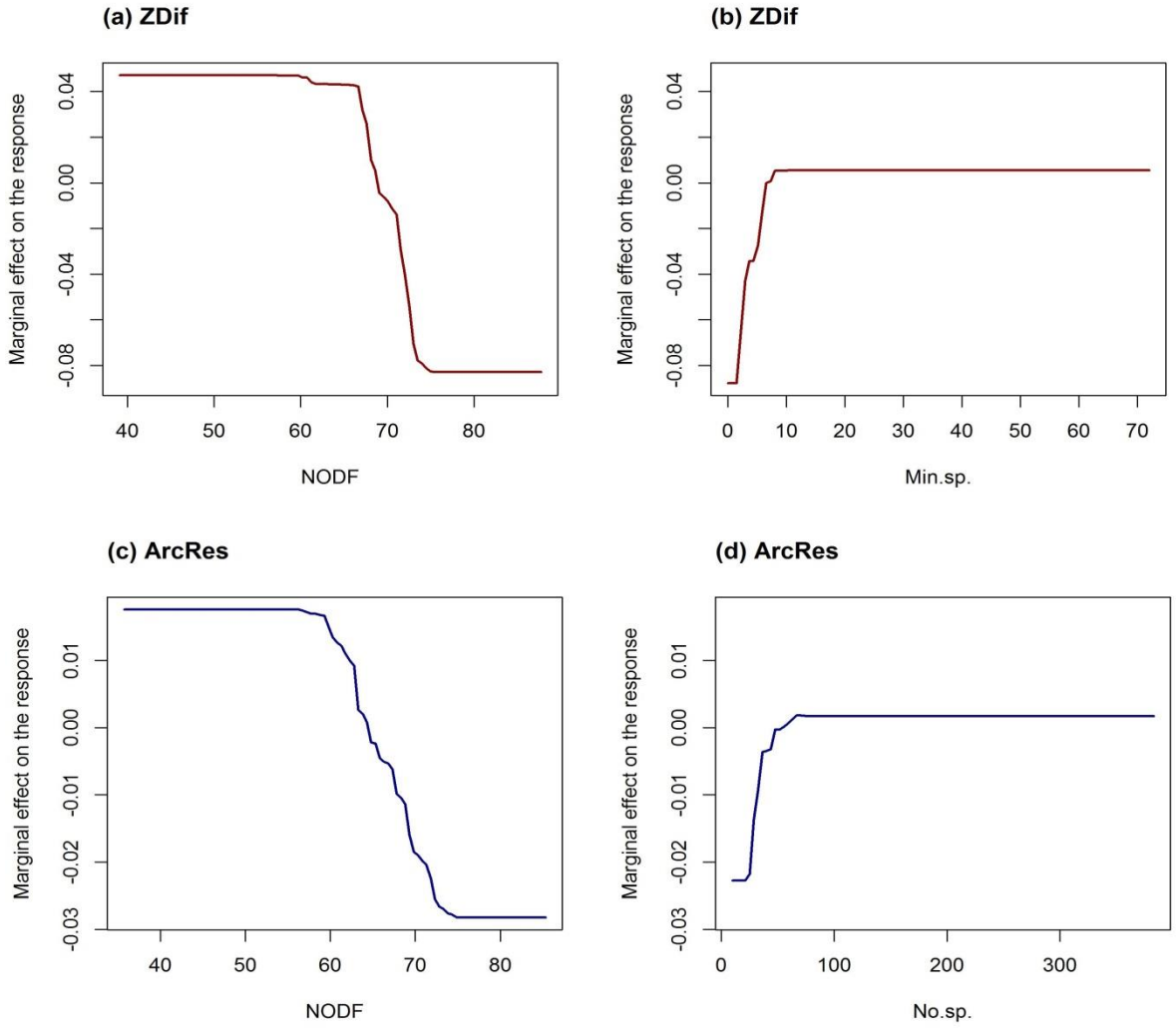


Figure 6.2 Partial dependence plots for a couple of the most influential variables in two boosted regression tree (BRT) models. The response variable was zDif (the difference between the z value of the SAC and the z value of the ISAR, for each dataset), and in the second model the response was ArcRes (c-d; the standardised difference between the cumulative species richness total of the archipelago and the predicted the number of species in the archipelagic point using the ISAR model of the constituent islands). Only zDif values from satisfactory datasets were used (n = 50) in (a) and (b), and only ArcRes values from datasets in which the fit of the power (log–log) model was significant (n = 73, reduced to 64 after removing nine outliers) were used in (c) and (d). The two predictor variables in the first model are (a) nestedness (NODF; for the relative contributions of each variable see Table 6.1) and (b) the minimum number of species in a dataset (Min. sp.). The two predictor variables in the second model are (c) nestedness (NODF), and (d) the number of species in a dataset (No. sp.). The y-axis shows the effect of a predictor variable (x-axis) on the response variable after accounting for the effects of the other model predictors (i.e. the marginal effect of the predictor). The y-axis represents the effect of the predictor on the response, and is not an indicator of the value of the response at a given value of the predictor. A positive y-value indicates that at the given x-value, the effect (based on the model) on the y-value (the response) is positive, and *vice versa*. The ends of the curves represent areas of the plotting space with fewer data points and are relatively uninformative.

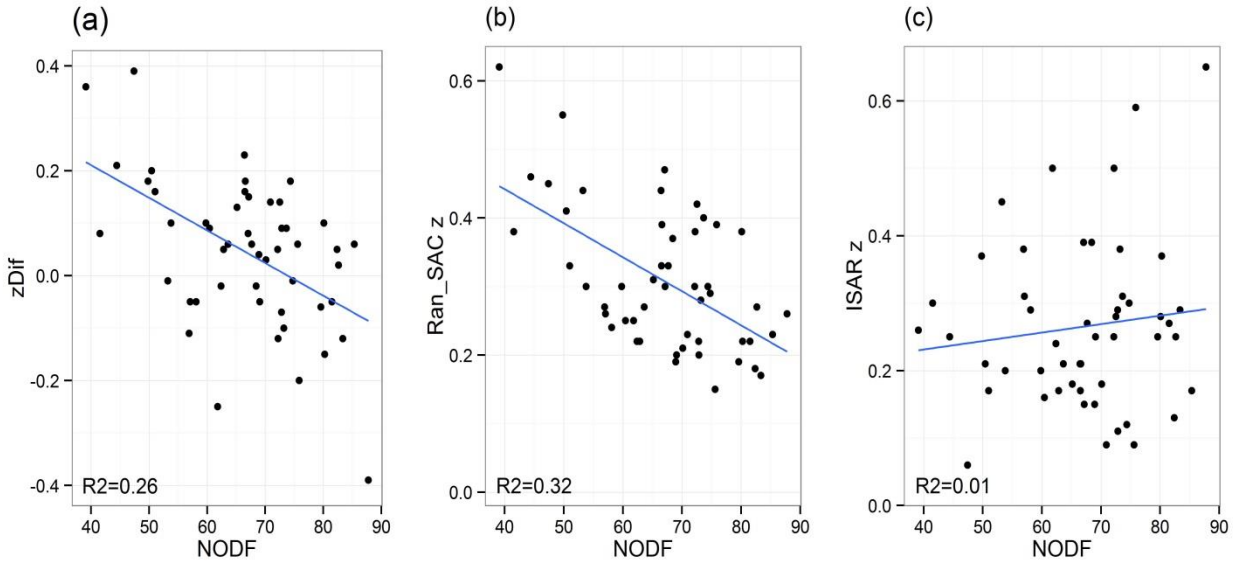


Figure 6.3 The relationship between nestedness (measured by the NODF metric) and (a) *z*Dif (the difference between the Ran SAC *z* value and the ISAR *z* value, (b) the SAC *z* value, and (c) the ISAR z value. Only *z* values from the satisfactory datasets (n=50) were used to construct the plots. In each plot, the solid line represents the best line of a linear regression model, and the R^2 of this fit is given on each plot. The slope was significant in each plot apart from (c), which is shown for illustrative value only. Increasing NODF value indicates an increasing degree of nestedness, according to this metric. The *z* values relate to the non-linear power SAR model.

6.5.3 Nestedness simulations

The simulation analyses revealed further evidence for the importance of nestedness in explaining the difference in *z* value between ISARs and SACs. In the case of Sim1, when the system was anti-nested (i.e. had a low NODF value), the slope of the SAC was greater than that of the ISAR. As the degree of nestedness was increased, the difference in *z* values declined until the slope of the ISAR became greater than that of the SAC (Fig. 6.4a). Figure 6.4b illustrates this effect for two small–large SAC curves (SL; the order of island incorporation into the SAC starts with the smallest island and increases up to the largest; see Appendix 4), constructed from the most nested and anti-nested Sim1 iterations, respectively.

We used the SL curve type in Fig. 6.4b as the area range of the ISAR and SAC is more similar and makes the plot easier to interpret.

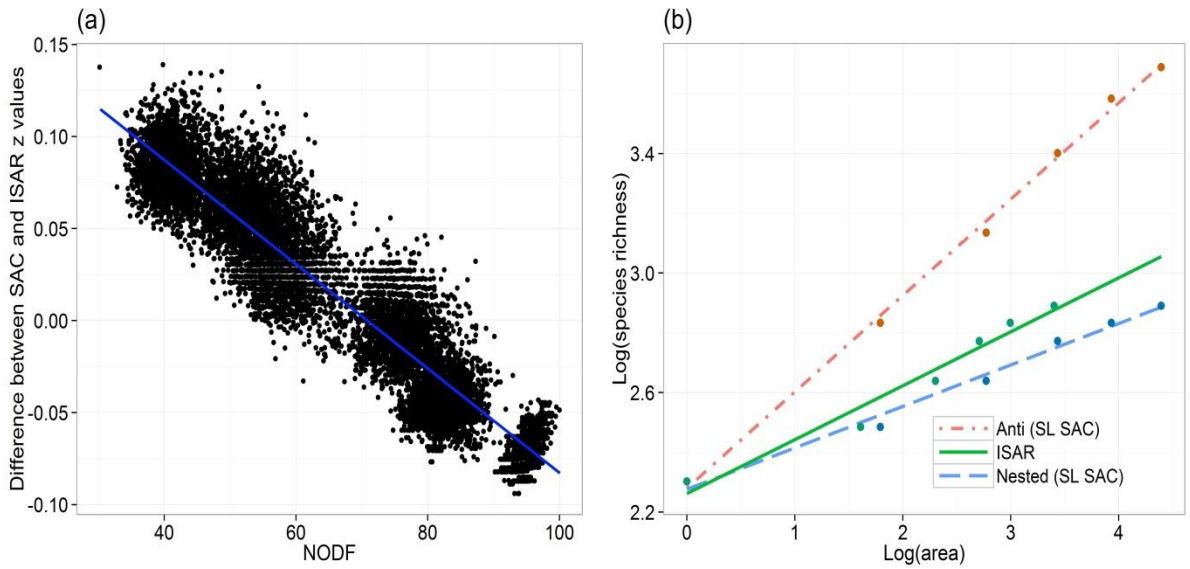


Figure 6.4 The relationship between nestedness and the variation in *z* values of ISARs and small-large (SL; the order of island incorporation into the SAC starts with the smallest island and increases up to the largest; see Appendix 4) constructed SACs. For (a), a perfectly nested presence/absence matrix for a set of six islands of varying area was simulated (area of each island =1, 5, 10, 15, 20 & 30; species richness of each island = 10, 12, 14, 16, 17 & 18). First, the power model (non-linear) was fitted to the data matrix in ISAR form, and the *z* value recorded. The SL SAC was then constructed using the same data matrix and, again the power model was fitted. The presence/absence matrix was then rearranged to change the level of nestedness, with the constraint that the species richness of each island was kept constant (i.e. the ISAR remained unchanged), although the overall number of species in the archipelago was allowed to vary; the power model was then fitted to both the ISAR and SAC constructed using this new data matrix. This process was repeated iteratively along a gradient of NODF (by rows, i.e. sites) values. A total of 12,000 matrix permutations were simulated. The best fit linear regression line (blue line) through these points (dots) is also shown. (b) The power (log-log) model, fitted to ISAR structured data (solid line, z = 0.18), and to SL SAC structured data using the same simulated island data as for (a). The fit of SL SAC curves to two matrix permutations are shown: a perfectly nested set of isolates (dashed line, z = 0.14), and a perfectly anti-nested set of isolates (dashed and dotted line, z = 0.32).

6.5.4 The fit of the archipelagic data point

Considering only datasets in which the power (log–log) model provided a significant fit (n=73), the archipelagic point deviated from the prediction of the constituent ISAR in 14 datasets (see Fig. 6.5 for an example of both scenarios). The ISAR under-predicted the archipelagic richness in 45 cases (Table S2). Considering 64 datasets (the 75 significant fits minus the nine outliers), when ArcRes was used as a response variable in a BRT analysis,

NODF was the variable with the highest relative influence value (Table 6.1). The number of species and the minimum number of species in a dataset were also important variables (Table 6.1). The mean coefficient of the correlation between the observed and predicted response values was 0.35.

6.6 Discussion

To evaluate the impact of confounding ISAR and SACs in SAR analyses and to understand their differing emergent properties we compared the slopes of ISAR and SAC curves using 97 habitat island datasets. We found that the slopes and form of the ISAR and SAC curves varied considerably within datasets, and that this variation was primarily explicable as a function of the degree of nestedness in the archipelago. We also observed that the ISAR was often a poor predictor of SAobs, demonstrating the dangers of extrapolating the ISAR to predict the richness of an archipelago of fragments.

6.6.1 Why are SACs generally steeper than ISARs?

In accordance with hypothesis 1 we have found often considerable differences between the slopes of ISARs and SACs (e.g. Fig. 6.1). Thus, our results illustrate that different inferences may be drawn depending on which SAR type is used in an analysis. In accordance with hypothesis 2, the difference in the slopes of SACs and ISARs is driven by nestedness. Intuitively it makes sense that nestedness drives the differences between the two curves (Gray *et al.,* 2004a, b); a fact highlighted by both empirical and simulation analyses (Figs 6.2, 6.4). The higher R^2 for the SAC compared to the ISAR (e.g. Fig. A4.1) also makes sense because the former represents a classic smoothing process in which richness must always remain constant or increase with area sampled, thus reducing scatter (Drakare *et al.*, 2006).

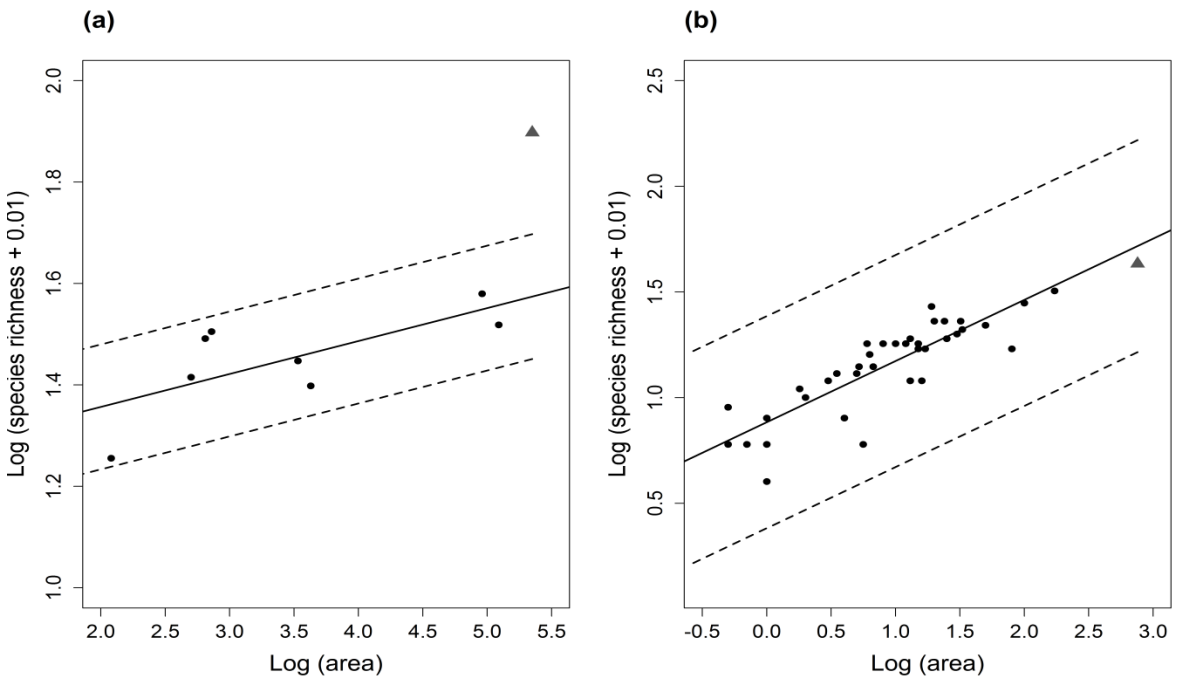


Figure 6.5 The island species–area relationship (power log–log model) for two habitat island datasets: (a) Benedick *et al.* (2006; invertebrates in Malaysia; number of islands = 8), and (b) this study (birds in the UK; number of islands = 37). The dots represent individual habitat islands and the fit of the power (log–log) model is indicated by the solid line. In both plots the archipelagic point, calculated as the total cumulative area and species richness of all habitat islands in the dataset, is illustrated as a triangle. The intervals defined by the maximum residual criterion are shown as the dashed line: this criterion relates only to the difference between the ISAR prediction and the archipelagic point, the dashed lines have then just been continued down the y-axis. If the archipelagic point lies outside these bounds (e.g. (a)) it is said to deviate from the prediction of the ISAR calculated using the constituent habitat islands; and *vice versa* (e.g. b; cf. Santos *et al.*, 2010).

A number of mechanisms have been argued to underpin the slope of SACs, through an effect on species overlap and/or nestedness (e.g. Quinn & Harrison, 1988; Chase & Knight, 2013). Three of these factors are particularly relevant to habitat island systems. First, a large species pool means that, all else being equal, there is a smaller probability that the same species will colonise an empty island first, and thus species overlap between islands is reduced. Furthermore, if islands draw species from different species pools the likelihood is that species overlap across all islands will be reduced (Whittaker & Fernández-Palacios, 2007). Second, if habitat types exhibit substantial turnover across islands, and species have strong affinity to particular habitat types, species overlap will be reduced (Quinn & Harrison,

1988). This is more likely to be an important factor if the islands are relatively farther apart due to the often reported distance decay in habitat types (Rosenzweig, 1995). Third, habitat islands are generally disturbed systems, and disturbance usually has a disproportionate effect in smaller islands (Whittaker & Fernández-Palacios, 2007). This means that matrix and generalist species are more likely to be present in smaller relative to larger islands (Matthews *et al.* 2014b), again resulting in reduced overlap. Each of these factors will vary in importance between habitat island systems and thus it is difficult to determine which factor is the most important. Nonetheless, in previous work we have shown that the relatively high incidence of generalist species in smaller fragments reduces the slope of a multimodel ISAR curve for a number of habitat island datasets (Matthews *et al.*, 2014a). This reduction in ISAR slope coupled with the reduced species overlap across islands provides one explanation for the many positive *z*Dif values we have observed.

Our analyses indicated that other mechanisms are also important, independent of nestedness (Table 6.1). For example, the minimum number of species (Min.) on an island in a dataset was the second most influential variable explaining variation in *z*Dif (Table 6.1). The partial dependence plot of this variable (Fig. 6.2b) indicates that low Min. values have a negative effect on *z*Dif; in this case, in many datasets with low Min. values the ISAR is steeper than the SAC (i.e. *z*Dif is negative). The reasons for this finding are unclear, but interestingly, the maximum number of species and the total number of species in the archipelago had low relative influence values (Table 6.1). Further research is needed to fully explore this result.

Landscape context variables (e.g. the matrix, and island type) were relatively unimportant (Table 6.1). This is surprising as the matrix is considered to be important in determining differences in slope between ISARs and SACs due to the influence of matrix properties on species turnover (see Crist & Veech, 2006; see Introduction). The lack of an

effect of these variables is likely due, at least in part, to the coarse matrix and island type classifications utilised in our analyses. Fine scale matrix properties are known to influence species distributions (Ricketts, 2001), but such fine scale data are rarely provided in the source papers.

6.6.2 Interpreting variation in ArcRes

Using the simple descriptive metric adopted by Santos *et al.* (2010) we found that the archipelagic point deviated substantially from the ISAR prediction in 19% of datasets with a significant ISAR, compared to 12% in Santos *et al.*'s analyses of true island datasets. NODF had the largest effect on ArcRes (Table 6.1), and this difference in outcome may reflect the fact that habitat islands often contain a large number of singletons (i.e. species that are only sampled in one habitat island in a dataset and whose presence reduces nestedness) as a result of factors such as source–sink dynamics and transient species. Furthermore, the ISAR more often under-predicted the richness of the archipelagic point (Fig. 6.5a; see Crist & Veech, 2006). Again, this implies low levels of nestedness in a number of systems (see also Matthews *et al., in press*) and brings into question the extrapolation of ISARs (see also Fig. 6.1b): a common applied use of the ISAR (Whittaker *et al.,* 2005). A criticism of the aforementioned Chong & Stohlgren (2007) extrapolation study, by Hui (2008), was that Chong & Stohlgren's extrapolations generally under-predicted the number of species present in different vegetation types precisely because they used a type of SAR more akin to an ISAR than a SAC. Our results are confirmatory of Hui's (2008) argument. It is important to highlight that this analysis contains the unstated assumption that SAobs equals the true total species richness. In reality, SAobs is likely an underestimate due to incomplete sampling. However, we have no reason to believe this has substantially affected our results.

6.6.3 Implications and conclusions

The choice of SAR type is an important consideration when using the SAR for applied purposes (Whittaker *et al.,* 2005), but conservation biogeographic studies are often unclear about the type of SAR employed (see discussion in Whittaker & Fernández-Palacios, 2007). In the introduction we highlighted a number of situations in which erroneous inferences could be drawn as a result of failing to distinguish ISARs and SACs (see also Gray *et al.,* 2004a, b). Our analyses have now shown for the first time using a large number of datasets why this is problematic. For instance, it is apparent that the ISAR slope is not related to the degree of nestedness (Fig. 6.3d), and thus slopes of the ISAR should not be used as indices of nestedness or community turnover (see also Wilson & Shmida, 1984; Gray *et al.,* 2004a). Arguably, this observation is not a surprise as logically it makes sense that slope of the SAC should be underpinned by the degree of nestedness, whilst the slope of the ISAR should be independent of nestedness. However, the distinction has not always been recognised. Thus, we call on researchers to explicitly identify the form of SAR used in future ecological analyses (see also Scheiner, 2003). We are not arguing that either type of SAR should be universally preferred; rather, the choice of which type of SAR to use depends on the aim of study and the data available. For example, SACs are more useful than ISARs for assessing the contribution of different sized islands to a protected area network (Fischer & Lindenmayer, 2002; Watson *et al.,* 2009). Instead, we are simply arguing that "the distinction between [island] species-area and species accumulation curves is extremely important" (Gray *et al.,* 2004a, p. 474).

Based on the self-similar scaling properties of the power model, Tjørve & Tjørve (2008; see also Harte *et al.,* 1999) used the SAC to show mathematically that the proportional overlap between two areas of the same size can be plotted as a function of *z*: as *z* increases, the degree of species overlap decreases. For example, a *z* of 0.58 equates to 50% of the

species being shared between the two areas (Tjørve & Tjørve, 2008). Our results provide an empirical illustration of this point (e.g. Fig. 6.3b). However, the nature of empirical data makes it harder to provide precise predictions/statements such as those in Tjørve & Tjørve (2008). Nonetheless, our results indicate that (a) the SAC is often steeper than the ISAR when the curves are constructed using habitat island data, and (b) the *z* value of the ISAR only becomes consistently steeper than the *z* of the SAC when the NODF value is approximately 70 or above (e.g. Fig. 6.3a & Fig. 6.4b).

It is also evident that for a number of habitat island systems the ISAR is a poor predictor of the overall number of species in an archipelago (e.g. Fig. 6.5a). Thus, caution should be employed when using the ISAR for extrapolation purposes in fragmented systems, particularly when species overlap is thought to be low (Crist & Veech, 2006). The SAC is likely to provide more accurate results in such contexts as it incorporates information on the degree of nestedness/overlap in the region (Quinn & Harrison, 1988; Hui, 2008). It is already acknowledged that the choice of sampling design is an important consideration in SAR studies, and our results provide more evidence for those who have recently argued for a more consistent SAR theory applicable to general SAR sampling designs (e.g. Whittaker *et al.*, 2005; He & Hubbell, 2013).

CHAPTER 7

Habitat fragmentation and the species–area relationship: a focus on total species richness obscures the impact of habitat loss on habitat specialists

Paper published in Diversity & Distributions: Matthews, T.J., Cottee-Jones, H.E.W. & Whittaker R.J. (2014) Habitat fragmentation and the species–area relationship: a focus on total species richness obscures the impact of habitat loss on habitat specialists. *Diversity and Distributions*, **20**, 1136–1146.

7.1 Preface

This chapter is the fourth empirical chapter, and examines the impact of including habitat generalist bird species in species–area relationship studies focused on habitat islands. Nineteen bird-habitat island datasets were sourced from the literature and all birds were classified as forest specialists or generalists. It was found that generalist species act to depress the slope of the ISAR, which can mask important patterns of decline in specialist species. The conservation biogeographic implications of these findings are discussed.

Authors: Thomas J. Matthews, H. Eden W. Cottee-Jones and Robert J. Whittaker

The co-author contributions were as follows. T.J.M. and R.J.W. conceived the ideas; T.J.M. and HEWC-J collected the data and classified the species; T.J.M. ran the analyses; and T.J.M. and R.J.W. wrote the paper with a significant contribution from HEWC-J. All authors discussed the results and commented on the manuscript.

7.2 Abstract

Aim The species–area relationship (SAR) is widely used in conservation science to predict the number of species likely to go extinct as a result of habitat loss. Often, studies employing the SAR use total species richness as the dependent variable. However, this overlooks the fact that habitat specialists and generalists differ in their susceptibility to habitat loss. We undertook a synthetic review of 23 habitat island datasets for birds to determine the impact of habitat generalists on the SAR.

Location Global

Methods We sourced 19 habitat island datasets from the literature and combined these data with four of our own empirically gathered datasets. For each dataset we classified all bird species as either forest habitat specialists or generalists. We then fitted the power SAR model (log-log and non-linear forms) to the specialists, generalists, and all species for each dataset and compared the resulting model parameters. We compared differences in the rate of change in richness with area between specialists and generalists using the first derivative of a multi-model SAR.

Results We found that the slope of the power model was steeper for habitat specialists in a majority of datasets, and this difference was significant in 15 and 16 out the 23 datasets, for the non-linear and log–log forms of the power model, respectively. Comparison of the multi-model SAR curve derivatives revealed further differences in the rate of change in species richness with area between subsets.

Main conclusions The *z* values of both forms of the power model of the specialists' SARs were generally larger, often considerably so, than the values used in most SAR studies predicting extinctions from habitat loss. Thus, studies that have used *z* values derived from SAR studies using total richness may be underestimating the impact of habitat loss on the loss of specialist species, which are likely to be those of greatest conservation concern.

7.3 Introduction

The increase in the number of species with the area sampled is one of the oldest and most reported patterns in ecology (e.g. Arrhenius, 1921; Rosenzweig, 1995; Lomolino, 2000; Whittaker & Fernández-Palacios, 2007). Various types of species–area relationship (hereafter SAR) have been documented for multiple taxa and been found to hold at a variety of spatial scales (Drakare *et al.,* 2006; Whittaker & Fernández-Palacios, 2007; Dengler, 2009). SARs have been widely used in biodiversity conservation to aid in the design of protected areas (e.g. Diamond, 1975a), the setting of conservation targets (Desmet & Cowling, 2004), and to predict extinctions resulting from habitat loss and fragmentation (e.g. Brooks *et al.,* 1997; Magura *et al.,* 2001; Drakare *et al.,* 2006; He & Hubbell, 2011). As the anthropogenic fragmentation and destruction of natural habitats is regarded as the primary causal factor driving the decline in global biodiversity (Wilcove, 1987; Whittaker & Fernández-Palacios, 2007), the development of the SAR as a predictive tool is of particular importance within conservation biogeography.

Herein we are concerned with type IV SARs sensu Scheiner (2003), also known as island species–area relationships or ISARs (Whittaker & Fernández-Palacios, 2007), which document the number of species within each isolate versus the area of each isolate. The majority of ISAR studies – including those concerning habitat fragmentation – use the total number of species recorded in the isolate as the dependent variable. A focus on total richness may be sub-optimal as it ignores the fact that some species are more susceptible to fragmentation than others because they differ in their ecological traits, habitat dependency and resilience to disturbance (Whitcomb *et al.,* 1981; Humphreys & Kitchener, 1982; Howe 1984; Henle *et al.,* 2004; Ewers & Didham, 2007). Viewing a fragmented landscape as a simplified binary of patches of natural habitat surrounded by a matrix of anthropogenic habitat, species can be dichotomised as habitat specialists and habitat generalists. Specialists

are those species acutely dependent on resources located in the patches and tend to be species of conservation concern (Lövei *et al.,* 2006). Other species recorded in habitat fragments persist and utilise resources in the surrounding matrix, and are, therefore, less dependent on the area of the patch ('habitat generalists'; e.g. Brotons *et al.,* 2003; Magura *et al.,* 2008; Ruiz-Gutiérrez *et al.,* 2010; Fahrig, 2013).

While the fragmentation of natural habitat often leads to the extirpation of habitat specialists, it can also result in an influx of generalist species, which acts to conceal the decline of the specialists. Hence, as the patch core:edge ratio declines with decreasing patch size the ratio of generalist to specialist species is likely to increase (Humphreys & Kitchener, 1982; Magura *et al.,* 2001). If habitat generalists are less dependent on habitat area, it follows that their inclusion in ISAR or other SAR calculations must affect the resulting model parameters (cf. the species accumulation curve analyses of Cook *et al.,* 2002). Thus, it has been suggested (e.g. Magura *et al.,* 2001; Bommarco *et al.,* 2010) that analyses should be based on subsets of species classified according to their level of specialisation.

If the depression of the ISAR slope through the inclusion of generalists is a general pattern it implies that the impacts of fragmentation are likely to be more deleterious than predicted by models based on total richness. Despite these implications, there have been few studies to date that have empirically investigated the impact of habitat generalists on ISAR parameters (but see Magura *et al.,* 2001, 2008; Watling & Donnelly, 2008; Bommarco *et al.,* 2010; Banks-Leite *et al.,* 2012), and we are unaware of any synthetic analysis of multiple datasets to determine the effect of generalist species on ISAR parameters.

Herein we undertake a synthetic review of habitat island datasets for forest birds to determine the impact of habitat generalists on the ISAR. For each dataset we classify species as either habitat generalists or specialists based on their level of forest dependency. We then use two analytical approaches to determine the impact of habitat generalists on the ISAR in

these datasets. First, we focus on the power SAR model (Arrhenius, 1921) as this is the most widely used model in conservation studies (Rosenzweig, 1995). The majority of studies utilise the linear (i.e. log–log) version of the power model (e.g. Watling & Donnelly, 2008). However, given modern advances in computing we can also fit the (intrinsically superior) non-linear version of the power model, which is particularly useful as the two versions result in different parameter estimates (Triantis *et al.,* 2012). Thus, we use both versions to ensure our results are relevant to as broad a range of studies as possible. We fit both versions of the model to the specialist and generalist species separately, for each dataset, and compare the parameter estimates using various methods. Second, we fit a set of eight candidate ISAR models and generate a weighted multi-model ISAR curve for the specialists and generalists separately, from which we compare the rates of change in species richness with area for each dataset. We focus on birds as the strongest evidence of the negative impacts of habitat fragmentation has come from studies on bird populations in forest fragments (Robinson & Sherry, 2012). Birds are also a well-studied taxon, possessing traits that enable them to be accurately classified in regards to habitat specialisation.

7.4 Methods

7.4.1 Data acquisition and species classification

Relevant studies were identified via an extensive search of the databases 'ISI Web of Science' and 'Scopus' between May 2011 and June 2013. A wide range of search strings were used, involving rearranging and amalgamating the phrases 'habitat fragmentation,' 'species–area relationships' and 'birds.' Each potential dataset was reviewed in order to ascertain that the following conditions were met:

1. A minimum of five forested habitat fragments had been sampled.

2. Each fragment conformed to Watson's (2002) definition of a true habitat island; that is, an area of natural habitat (in this case forest) surrounded by a matrix of a contrasting land use type.

3. The area of each fragment was presented.

4. The studies included a full inventory of the bird species in each fragment.

5. The dataset did not overlap with that from any other study that had been accepted for analysis.

In addition to the above, four fragmented landscapes in Europe (France, Spain, Norway & the UK) were sampled by TJM and the resulting datasets added to the analysis. In each of these landscapes around 40 forest fragments of varying area (see Appendix 5 in Supporting Information for details) were sampled using ten-minute point counts of 50m radius. All birds seen or heard were recorded. Counts took place from dawn to 10am during April–July in 2012 and 2013. Each fragment was sampled three times, with the time of sampling randomised in each instance.

To determine the habitat specialisation level TJM and HEWC-J classified each species independently based on a large amount of material sourced from the ornithological and wider ecological literature (a full list of the references and key terms used in the literature searches is provided in Appendix 5). We then compared our independent classifications, and undertook further investigation (including obtaining additional opinions from various senior ornithologists) to classify species on which we initially disagreed. Species were classified based on their level of habitat specialisation in regards to the habitat island type in each particular study (i.e. classification was not based on any metric of sensitivity to patch area).

This classification was determined using a selection of information regarding species' traits, and information on where the species had previously been recorded.

7.4.2 Data analysis

Three versions of each dataset were created: area – specialist species richness (herein 'SAR_{sp}'), area – generalist species richness (herein 'SAR_g'), and area – total species richness (herein 'SAR_t'). Firstly, we fitted the linear version of the power function (log–log model) of Arrhenius (1921) to each version of the datasets, after log-transforming both variables, using linear regression:

$$LogS = Log\mathit{c} + zLogA \qquad (1)$$

where S=species richness, A=area and *z* and *c* are fitted constants. All fragment areas were converted to hectares to permit comparison of the *c* parameter (Rosenzweig, 1995). All logarithmic transformations were to the base ten. As log(0) is non-defined, log(x+0.1) was applied for the transformation of species richness in each case. We then fitted the non-linear power model to each version of the datasets using non-linear regression and the 'mmSAR' R package (Guilhaumon *et al.*, 2010):

$$S = \mathit{c} \,.\, A^{z} \qquad (2)$$

To determine ISAR significance for the power models, the slope of the observed regression line was tested in each case against a null hypothesis of the slope being equal to zero ($P < 0.05$). To determine whether the parameters (*z* and *c*) of the log–log power model significantly differed between different versions of each dataset, ANCOVA was used to compare the regression lines of SAR_{sp} and SAR_g. As SAR_{sp} and SAR_g are not independent of SAR_t, we could only compare the regression lines of SAR_{sp} and SAR_g using the ANCOVA model. A critical value of $P < 0.05$ was used. In addition, we calculated the effect size (Cohen's D) for each dataset using the ANCOVA F score and the 'compute.es' R package (Del Re, 2010). We then calculated the pooled effect size (d+) using the weighted mean of

the individual effect sizes, where the weights represented the inverse of the variance of the individual effect sizes (Hedges & Olkin, 1985). A Z score and *P* value were generated to determine whether d+ was significantly different from zero.

Using ANCOVA is not possible for non-linear models. Thus, to compare the *z* value of the non-linear power model between specialists and generalists we generated 95% confidence intervals for each parameter value and the *z* values were said to differ if the confidence intervals did not overlap (cf. Gaston *et al.,* 2006).

While the power model is the most frequently used ISAR model, it should not be assumed that it provides a better fit than competitor SAR models *a priori* (although for real islands, a recent synthetic analysis shows that it is the best performing model overall: Triantis *et al.,* 2012). Thus, we fitted a set of eight SAR models (power, exponential, negative exponential, Monod, logistic, ratio, Weibull-3, Lomolino) to the SAR_{sp} and SAR_g versions of each dataset using non-linear regression and the mmSAR R package. Models were compared using Akaike's information criterion corrected for small sample size (AIC_c; Burnham & Anderson, 2002). Models with the lowest AIC_c value were considered to provide the best fit, and all models within Δ AIC_c <2 of the best model were considered to have similar support (Burnham & Anderson, 2002).

We used the 'multiSAR' function in the mmSAR package to generate a multi-model SAR curve for each version of the datasets by weighting the predicted values of the eight aforementioned SAR models based on the Akaike weights of the individual models (Guilhaumon *et al.,* 2010). In order to compare the multi-model SAR curves between SAR_{sp} and SAR_g for each dataset we calculated the first derivative of the multi-model SAR curves (i.e. the rate of change in species richness with respect to area; cf. Diouf *et al.,* 2009) and plotted the derivative as a function of island area (i.e. we constructed a growth rate curve). This was done separately for each dataset and examination of the relative position of the

specialist and generalist curves in these plots allowed us to infer differences in the rate of change in species richness with area between specialists and generalists. As a further test we subtracted the generalist curve from the specialist curve: unless the curves cross, a positive result indicates that the specialist curve is steeper.

Following previous workers (e.g. Humphreys & Kitchener, 1982; Sfenthourakis & Triantis, 2009), we calculated the ratio of specialist to generalist species (herein 'S:G') in each habitat island and plotted this against island area for each study system. We also conducted boundary tests using EcoSim (Gotelli & Entsminger, 2001b) for each ratio plot to determine whether any corners of the plots contained fewer data points than expected by chance. Both the 'number of points' and 'sum of squares' boundary test methods were used. With the exception of the boundary tests, all analyses were conducted in R (version 3.0.0.; R Development Core Team, 2013).

7.5 Results

Over 250 published articles were screened, from which 19 datasets were deemed suitable for analysis (see Appendix 5 for details). Grouping these with our own four sampled datasets resulted in 23 datasets in total. All datasets comprised forested habitat islands surrounded by a matrix of contrasting habitat; in the majority of cases the matrix comprised agricultural land uses.

7.5.1 Power (log–log)

For 16 of the 23 (70%) datasets the slopes of the linear regression lines significantly differed between SAR_g and SAR_{sp} (Table 7.1), with the slopes (*z* values) being steeper for SAR_{sp} ISARs in each case (Table 7.1, see Fig. 7.1 for examples). Indeed, with the exception of system 22, which failed to provide a significant ISAR for SAR_{sp}, SAR_g, or SAR_t, all *z* values were higher for SAR_{sp} than SAR_g. For six of the seven datasets in which the slopes did not significantly differ, the intercepts were significantly different (results not presented). Only for

system 10 did both the slopes and intercepts of the specialist and generalist ISARs not differ significantly. The effect size results followed the ANCOVA results; that is, the datasets with significant values of Cohen's D were the same datasets in which the ANCOVA results were significant (individual Cohen's D values are given in Table 7.1). The pooled effect size (d+) was 0.53 (CIs: 0.39–0.66; Z=7.5, *P*=<0.0001), which indicates a medium to large effect size.

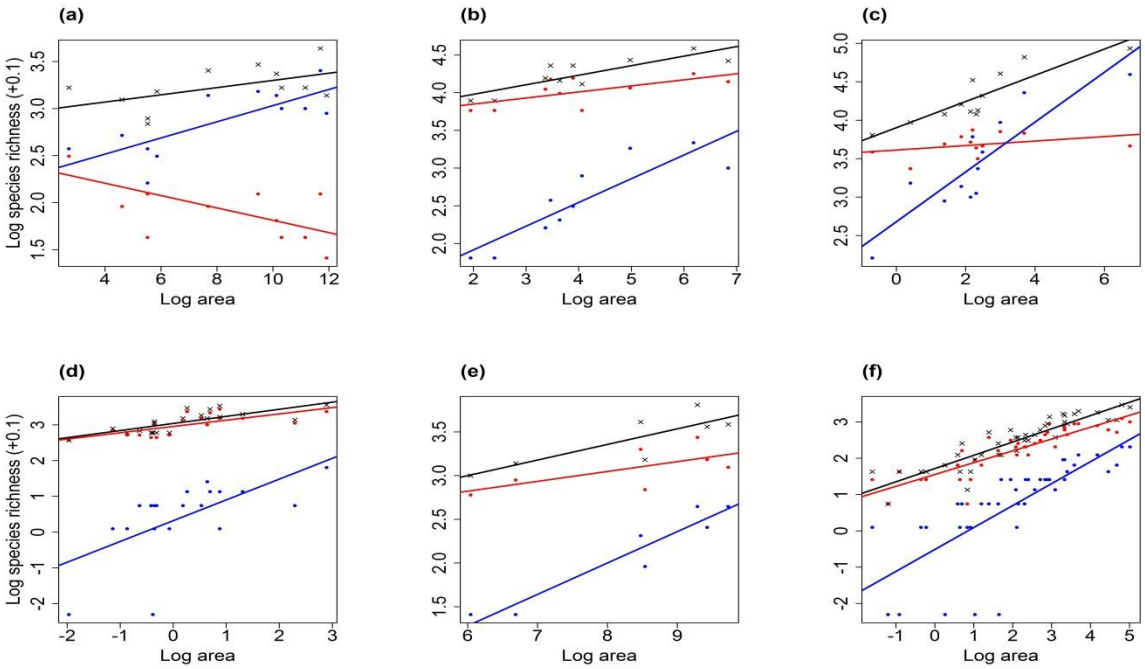


Figure 7.1 A selection of plots of the type IV SARs sensu Scheiner (2003) (here termed ISARs) using log–log power models, for six bird datasets. In each plot the data points (symbols) and linear regression best fit lines (solid lines) for all species (black crosses and line), specialists (blue dots and line) and generalists (red dots and line) are presented. In two instances the slope of the regression line for the generalist species was not significantly different from zero (c & e); in all other instances the slope of the regression line was significantly different from zero (i.e. $P < 0.05$). The sources of the datasets are: a) Anciães & Marini (2000), b) Castelletta et al. (2005), c) dos Anjos & Boçon (1999), d) Ford (1987), e) Gillespie & Walter (2001) and f) This study – Norway. See Appendix 5 for full details for each of the datasets. All areas are in hectares (log transformed to the base ten).

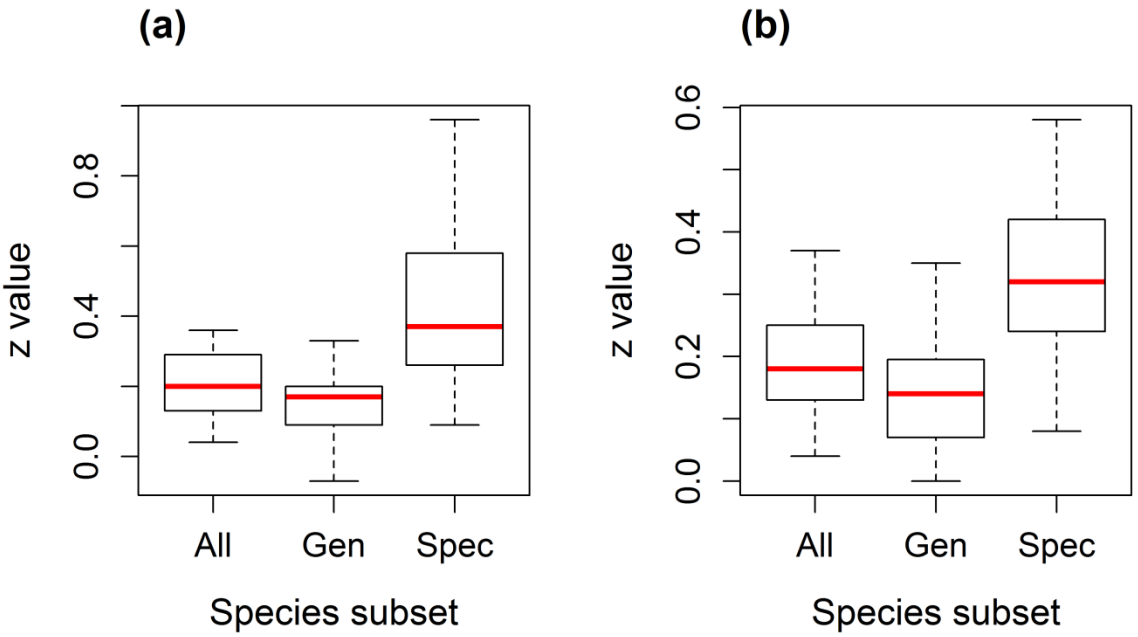


Figure 7.2 Box plots displaying the *z* values of the power model (log–log)(a) and the non-linear power model (b). Only *z* values from significant ISARs are used to generate the plots. *z* values are presented for the model fitted using all species in the dataset (all) and using generalist (Gen) and specialist (Spec) species separately.

7.5.2 Power (Non-linear)

While the parameter values generated from the non-linear power model (Table 7.2 & Table A5.2 in Appendix 5) differed from those generated by the log–log model (compare Table 7.1 with Table 7.2) the overall results were qualitatively similar. For 15 (65%) datasets the *z* value of the SAR_{sp} was greater than that of the SAR_g and the confidence intervals did not overlap (Table 7.2). For the remaining datasets, the *z* value of the SAR_{sp} was larger, but the confidence intervals did overlap.

A large range in *z* values was observed for SAR_t, SAR_{sp} and for SAR_g for both the log–log and non-linear versions of the power model (Fig. 7.2). Across all studies, the *z* values of SAR_{sp} and SAR_g differed significantly according to a Welch two sample t-test for both the log–log model (t = 3.34, P = < 0.001) and the non-linear model (t = 4.21, P = < 0.001). The *c* parameter was lower for SAR_{sp} compared to SAR_g and SAR_t, for all 23 systems, for both

forms of the model, and the difference in *c* values between SAR_{sp} and SAR_g was significant for both models (log–log: t = -4.86, *P* = < 0.001; non-linear: t = -3.72, *P* = < 0.001).

7.5.3 Multi-model inference

The power model (non-linear) was selected as the best model, or was within Δ <2 AIC_c of the best model, for 15 SAR_g datasets, and 14 SAR_{sp} datasets (Table A5.3 in Appendix 5). For the remaining datasets a variety of the other candidate models provided the best fit (Table A5.3). Analysis of the derivative plots and subtraction results indicated that the rate of change in richness with area was greater for specialists than generalists for 14 of the 23 datasets (see Fig. 7.3 for examples; all plots presented in Fig. A5.1 in Appendix 5). In one dataset (plot k in Fig. A5.1) the curves were almost identical, and in the remaining seven the generalists' curve was steeper. In terms of overlap with the power model results, in three datasets with a significant ANCOVA result the rate of change in richness with area for specialists was less than for generalists; and for four datasets with a non-significant ANCOVA result the specialists' derivative curve was steeper.

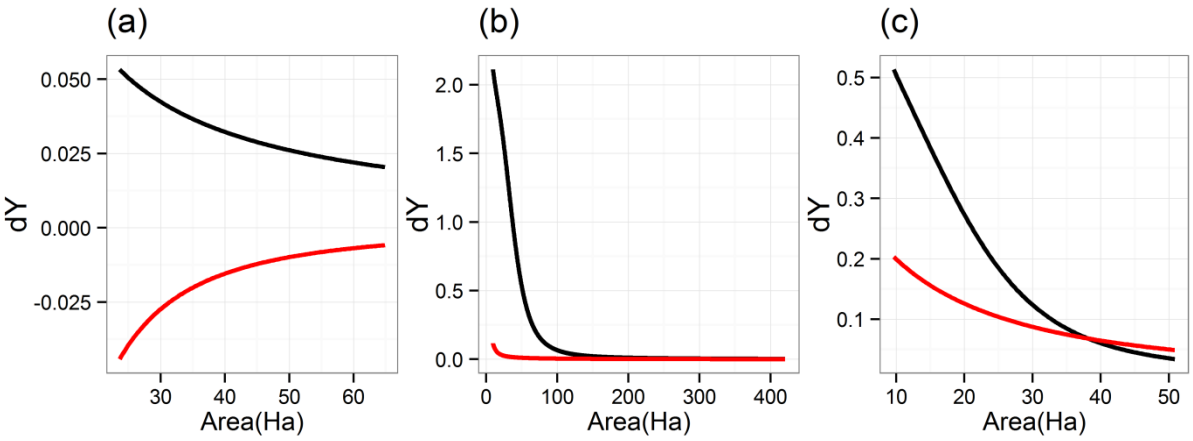


Figure 7.3 A selection of derivative plots for three bird-habitat island datasets. In each plot the curves represent the first derivative of the multi-model SAR curve (dY) plotted against island area, for both specialists (black line) and generalists (red line). Thus the curves represent the change in the slope of the multi-model SAR curve with increasing area. The plots have been truncated at the right hand side to improve presentation. The sources of the datasets are: a) Anciães & Marini (2000), b) dos Anjos & Boçon (1999), and c) This study – Norway.

Table 7.1 Model fit and ANCOVA results for the 23 bird–habitat island datasets. The *z* value of the power SAR function (log–log) model along with the *P* value corresponding to the significance of the *z* value in parentheses is presented. The *z* value is given for the model parameters calculated using all species, and specialist and generalist species separately. For each dataset the F statistic and corresponding *P* value of the ANCOVA model with species richness as the dependent variable, island area as the covariate, and species type (i.e. generalist or specialist) as the two level factor, is also given. Cohen's D effect size metric is also given in each case (calculated using the ANCOVA F statistic). All ANCOVA and Cohen's D *P* values significant at the 0.05 level are highlighted in bold.

	z (*P* value)			ANCOVA		
Dataset	All species	Specialists	Generalists	F Value	*P* Value	Cohen's D (*P*)
1	0.04(0.08)	0.09(0.00)	-0.07(0.04)	18.17	**0.00**	1.63 **(<0.001)**
2	0.31(0.00)	0.38(0.00)	0.20(0.00)	14.59	**0.00**	1.08 **(<0.001)**
3	0.13(0.00)	0.31(0.00)	0.08(0.04)	11.67	**0.00**	1.24 **(0.01)**
4	0.68(0.00)	0.95(0.00)	0.61(0.00)	11.43	**0.00**	0.28 **(<0.001)**
5	0.17(0.00)	0.32(0.00)	0.03(0.26)	23.54	**0.00**	1.45 **(<0.001)**
6	0.20(0.00)	0.58(0.00)	0.18(0.00)	5.34	**0.03**	0.62 **(0.03)**
7	0.68(0.00)	0.63(0.00)	0.62(0.00)	0.01	0.91	0.03 (0.92)
8	0.18(0.02)	0.36(0.00)	0.11(0.11)	10.62	**0.01**	1.21 **(0.02)**
9	0.11(0.04)	0.26(0.09)	0.08(0.20)	1.34	0.26	0.44 (0.26)
10	0.08(0.00)	0.09(0.00)	0.06(0.02)	1.23	0.28	0.32 (0.28)
11	0.15(0.00)	0.18(0.00)	0.13(0.00)	1.90	0.19	0.48 (0.2)
12	0.20(0.00)	0.29(0.00)	0.14(0.00)	7.81	**0.02**	0.35 **(0.04)**
13	0.25(0.00)	0.53(0.00)	0.20(0.00)	9.02	**0.00**	0.47 **(<0.001)**
14	0.36(0.00)	0.60(0.00)	0.33(0.00)	9.69	**0.00**	0.45 **(<0.001)**
15	0.19(0.00)	0.58(0.00)	0.17(0.00)	9.52	**0.00**	0.45 **(<0.001)**
16	0.29(0.00)	0.61(0.00)	0.25(0.00)	9.55	**0.00**	0.52 **(<0.001)**
17	0.20(0.00)	0.58(0.02)	0.17(0.01)	3.96	**0.05**	0.59 **(0.05)**
18	0.21(0.04)	0.38(0.01)	0.06(0.58)	9.67	**0.02**	1.42 **(0.04)**
19	0.62(0.00)	0.96(0.00)	0.55(0.00)	7.15	**0.01**	0.44 **(0.01)**
20	0.21(0.00)	0.32(0.00)	0.14(0.00)	4.96	**0.03**	0.52 **(0.04)**
21	0.07(0.02)	0.09(0.00)	0.06(0.06)	0.60	0.45	0.21 (0.45)
22	0.03(0.08)	0.02(0.26)	0.04(0.06)	0.78	0.39	0.32 (0.4)
23	0.12(0.00)	0.17(0.00)	0.09(0.02)	2.83	0.11	0.2 (0.12)

Table 7.2 The *z* parameter of the non-linear power SAR model and associated lower (LCI) and upper (UCI) 95% confidence intervals for the 23 bird–habitat island datasets. The *z* value is given for the model parameters calculated using specialist and generalist species separately. For each dataset the *z* values were said to be significantly different between generalists and specialists if the confidence intervals for the *z* parameter did not overlap. Specialist *z* values which are significantly larger than generalist *z* values are highlighted in bold.

	Generalists			Specialists		
Dataset	*z*	LCI	UCI	*z*	LCI	UCI
1	-0.07 (1.00)	-0.14	-0.01	**0.08(0.01)**	0.03	0.14
2	0.18(0.00)	0.12	0.22	**0.29(0.00)**	0.23	0.36
3	0.07(0.04)	0.00	0.14	**0.29(0.01)**	0.16	0.44
4	0.47(0.00)	0.42	0.51	**0.58(0.00)**	0.52	0.66
5	0.02(0.32)	-0.03	0.07	**0.25(0.00)**	0.17	0.34
6	0.15(0.00)	0.07	0.21	**0.36(0.00)**	0.22	0.53
7	0.16(0.00)	0.08	0.23	**0.39(0.00)**	0.24	0.54
8	0.11(0.16)	-0.02	0.22	**0.38(0.01)**	0.24	0.54
9	0.11(0.13)	-0.03	0.27	0.16(0.28)	-0.08	0.47
10	0.06(0.02)	0.01	0.11	0.09(0.00)	0.05	0.14
11	0.12(0.00)	0.07	0.17	0.15(0.00)	0.10	0.20
12	0.15(0.00)	0.09	0.17	**0.32(0.01)**	0.21	0.52
13	0.24(0.00)	0.18	0.30	**0.45(0.00)**	0.35	0.54
14	0.28(0.00)	0.23	0.33	**0.43(0.00)**	0.36	0.52
15	0.22(0.00)	0.16	0.28	**0.44(0.00)**	0.30	0.60
16	0.21(0.00)	0.16	0.27	**0.42(0.00)**	0.32	0.52
17	0.16(0.02)	0.04	0.29	0.3(0.04)	0.05	0.58
18	0.04(0.64)	-0.24	0.32	0.35(0.01)	0.16	0.58
19	0.35(0.00)	0.32	0.38	**0.43(0.00)**	0.40	0.43
20	0.14(0.00)	0.09	0.17	**0.24(0.00)**	0.21	0.33
21	0.07(0.07)	-0.01	0.15	0.08(0.01)	0.03	0.13
22	0.05(0.05)	0.00	0.09	0.02(0.26)	-0.02	0.06
23	0.1(0.00)	0.05	0.15	0.17(0.00)	0.11	0.24

7.5.4 Specialist:generalist ratio

Inspection of the plots of S:G against island area indicates that in the majority of cases S:G increases with island area but in a non-linear fashion; that is, after a certain island area the relationship tends to flatten out (see Fig. 7.4 for examples; all plots are presented in Fig. A5.2 in Appendix 5). The boundary tests revealed no significant results when using the ‘number of

points method' for any quadrant of the plots, and only one marginally significant result for the lower right quadrant (P=0.05; dataset 14) when using the 'sum of squares' method (results not presented). Thus, in all but one case there were no regions of the ratio plots that were unusually empty.

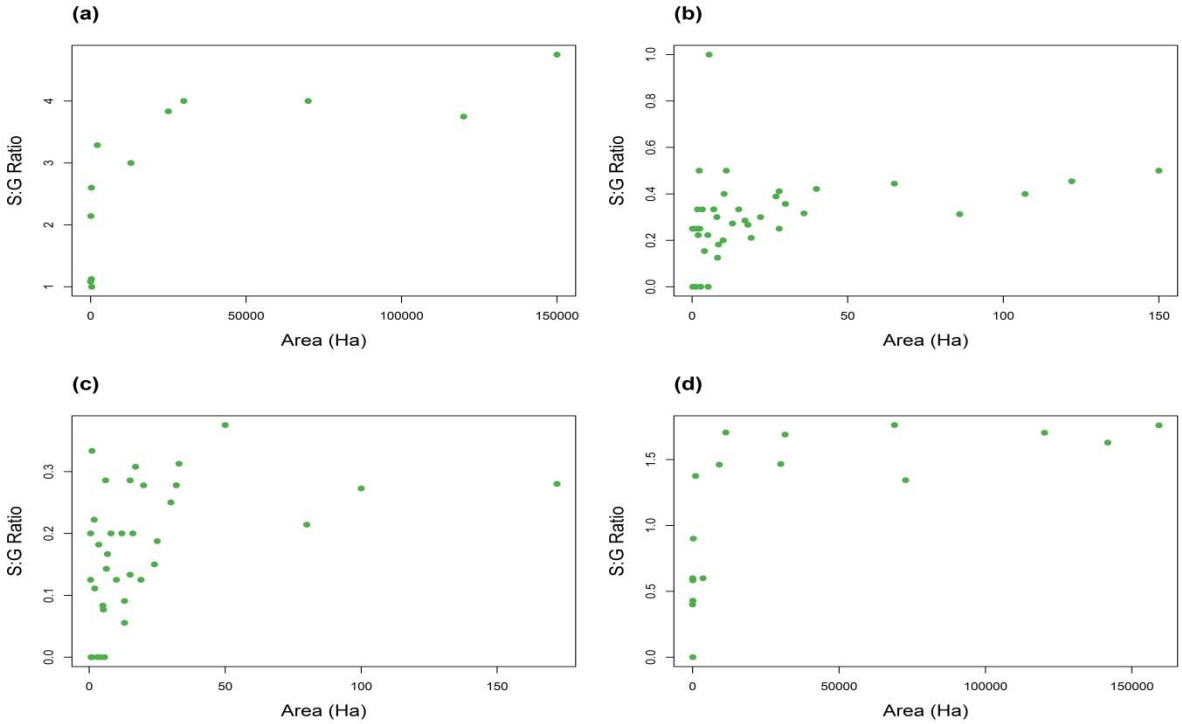


Figure 7.4 A selection of plots in which the ratio of specialist species to generalist species (S:G) against habitat island area is plotted for a selection of bird datasets. The sources of the datasets are: a) Anciães & Marini (2000), b) This study – Norway, c) This study – UK, d) Watson (2003). See Table A5.1 in Appendix 5 for full dataset information. Each of the plots illustrates the threshold pattern whereby over small areas the ratio of specialist to generalists increases rapidly, but after a point the relationship flattens.

7.6 Discussion

7.6.1 Generalist species mask the decline in specialists following habitat loss

The ISAR is an important conservation tool and is often used to predict the number of extinctions resulting from the loss of native habitat (see Fattorini & Borges, 2012). However, when calculating the ISAR total species richness is often used, which amalgamates habitat specialists and generalists into a single sample. We amassed 23 habitat island datasets for

forest birds and classified all species into generalists and specialists. We found that in 15 and 16 cases the slope of the specialists' ISAR was significantly steeper than the generalists' ISAR, for the non-linear and log–log versions of the power model respectively. Analysis of the multi-model derivative plots also reveals differences in the rate of change in species richness with area between subsets.

In predictive ISAR studies it is often the case that a set *z* value derived from the literature is used, such as the canonical value (Preston, 1962) of approximately 0.25 (e.g. Brooks *et al.*, 1997). However, our results show that the *z* value of SAR_{sp} was greater than 0.25 in 17 (74%) of the datasets when the log–log model was used, and in 14 (61%) of the datasets using the non-linear model (Table 7.1 and Table 7.2). In fact, the mean *z* value for SAR_{sp} was considerably larger than 0.25 for both models (Fig. 7.2). As specialists are largely the species that are at risk of extinction due to habitat loss (Howe, 1984), it seems unwise to assume a *z* value lower than the average for such species. These differences are not trivial. For example, consider a hypothetical 100ha patch of habitat containing 350 species which is reduced in area by 50%. Using the backwards version of power model (see Brooks *et al.*, 1997; Whittaker & Matthews, 2014) with a traditional *z* value of 0.25 results in 56 predicted patch extinctions, while a *z* of 0.95 (the highest we recorded) results in 170 predicted patch extinctions. This issue can be further exemplified by that the fact that in one dataset the ISAR (log–log) of generalists is in fact negative (Fig. 7.1a). Closer inspection of the species in this dataset reveals there to be a small number of species in the smaller patches which are primarily matrix specialist species (e.g. *Thraupis sayaca*), whose incidence declines with increasing area, contributing to the observed negative ISAR.

We are not the first to report *z* values for habitat islands exceeding the commonly used 0.25 value. Whereas some studies have argued that standard SARs always overestimate the number of extinctions (e.g. He & Hubbell, 2011), a number of studies have argued that

they usually underestimate extinctions (Brooks *et al.*, 2011; Fattorini & Borges, 2012). Such differences of view may reflect factors such as analytical approach taken, the approach taken to considering delayed extinctions (i.e. extinction debt; Triantis *et al.*, 2010; Tanentzap *et al.*, 2012), or the effects of fragmentation that generally accompany habitat loss (Fattorini & Borges, 2012). We caution, however, that in their analysis of 465 'real' island datasets, Triantis *et al.* (2012) observe strong scale dependency of the *z* value, which tends to decline with increasing range of island areas. Given that only four of our datasets have a maximum island size >1000 ha, two of which have *z* values lower than 0.25, it would be unwise to assume that the distribution of *z* values reported in Table 7.1 is representative of larger scales of application (such as e.g. the analyses of Brooks *et al.*, 1997). Although it is worth bearing in mind that the majority of habitat island datasets contain relatively small islands in relation to oceanic island datasets. These observations point to the need for further work to establish appropriate *z* values for applications involving extinction risk estimates for habitat-specialists and specifically for potential scale (range of area, range of isolation) sensitivity of ISAR parameters for habitat island systems.

Interpreting patterns in the *c* parameter of the power model is even less straightforward than for the *z* parameter (Gould, 1979; Whittaker & Fernández-Palacios, 2007). One commonly accepted interpretation is that *c* represents a measure of the carrying capacity of the system; that is, the species richness present at one unit of area (Gould, 1979; Rosenzweig, 1995). Thus, as the *c* value of specialists was always lower than for generalists, our results indicate that one unit area of habitat island forest can support fewer specialist species than generalist species. This is consistent with previous studies that have reported that, in general, habitat specialists require more patch resources than generalists to complete their life history (see Henle *et al.*, 2004), and that generalists are less reliant on patch resources as they are more able to utilise resources in the surrounding matrix (Brotons *et al.*,

2003). This pattern in *c* values again implies that habitat specialists are more sensitive to habitat loss than habitat generalists; combining the two sets of species acts to average out the parameter value in the aggregate.

The good performance of the power model for the majority of datasets is reassuring; given that it is the most widely used ISAR model (see also Triantis *et al.*, 2012). However, with the exception of the logistic model, every ISAR model was selected as the best model for at least one version of a dataset (Table A5.3); thus emphasising the potential relevance of considering multiple models in ISAR analyses. In this regard, our method of comparing the derivative curves is a practical and robust way of combining an assemblage deconstruction approach with multi-model inference, which is independent of species richness, and we recommend its use in future studies of this nature (see also Diouf *et al.,* 2009).

This is not the first study to advocate the use of a dependent variable other than total species richness when using the ISAR or other SARs for conservation purposes (e.g. Humphreys & Kitchener, 1982; Azeria *et al.*, 2007; Watling & Donnelly, 2008; Bommarco *et al.*, 2010). For example, a number of studies have advised using only endemic species and the related endemics–area relationship ('EAR'; e.g. He & Hubbell, 2011; Tanentzap *et al.*, 2012), and others have shown differences in ISARs between native and introduced species (Magura *et al.,* 2008). However, this is, to our knowledge, the first study to use multiple datasets to quantify how subdivision into generalists and specialists alters the parameters of ISARs; parameters which in turn are often used to guide conservation efforts (but see Bommarco *et al.,* 2010; Banks-Leite *et al.,* 2012 for examples using data from individual systems).

Our primary finding that the SAR_{sp} was steeper than the SAR_{g} was not universal and a number of datasets exhibited the opposite pattern. A possible explanation for this counter-intuitive result is that for these datasets a large number of specialist species have already been extirpated from the larger habitat islands, which acts to depress the SAR_{sp} (see Bommarco *et*

al., 2010). For example, the Castelletta *et al.,* (2005; dataset number 3 in Table A5.1) dataset is of birds in forest fragments on the island of Singapore. Singapore has suffered almost complete deforestation (approximately 99.5% of native forest has been removed) and habitat degradation, which has had significant negative effects on the native avifauna (Castelletta *et al.,* 2005). It is thus likely that the most patch-area-sensitive species have already been extirpated from the system, confounding our ISAR analyses. It is also possible that other life history traits, such as body size and dispersal capability (Ewers & Didham, 2007; Bommarco *et al.,* 2010), are more important than habitat specialisation in moderating the ISAR for species in these datasets.

7.6.2 Ratio of specialists to generalists

The ratio of specialists to generalists (S:G) generally increases steeply with patch area before appearing to flatten out (Fig. 7.4). Other studies have shown that S:G increases with area (e.g. Whitcomb *et al.,* 1981; Humphreys & Kitchener, 1982), but the flattening of the relationship has not been generally recognised. The initial increase in S:G with area results from the fact that certain species respond positively to fragmentation (i.e. edge and generalist species) and are thus more likely to be present in smaller patches (McCollin, 1993). By definition generalists have a wider niche and less specialised resource requirements than specialists, in addition to being better able to utilise resources outside the patch (Laurance & Bierregaard, 1996). Taken together these traits allow generalists to maintain high levels of occupancy in small patches. In contrast, specialists are expected to have a low incidence in smaller patches as both the resource base (Dennis *et al.,* 2012) and habitat diversity (Ricklefs & Lovette, 1999) are reduced. At large patch sizes the increased resource availability allows larger populations of specialists to coexist and outcompete the generalist species (Diamond, 1979). Thus, generalists and specialists have the opposite incidence function with respect to area, which explains the initial increase in S:G with area. The flattening of the S:G relationship

could be due to two reasons. First, the relationship may simply mirror the power law ISAR curve whereby in larger patches few species of any type are added. Second, as patch area increases richness continues to rise but specialists and generalists are added at a constant rate. Examination of the data suggests both processes are occurring. In certain datasets the S:G relationship flattens as the rate of species addition does decrease with area in accordance with a power law relationship, while in other datasets richness is still increasing in larger patches, albeit at a slightly reduced rate, but specialists and generalists are added at the same rate.

7.6.3 Conservation implications

Our results indicate that SAR_{sp} ISARs have higher *z* values than do SAR_g ISARs: in the majority of cases significantly so. Analysis of multi-model derivative curves provides further evidence for differences in the rate of change of richness with area between subsets. Thus, studies that have used total species richness, or *z* values derived from ISAR studies using total richness, to estimate the effects of habitat loss, may in this respect be underestimating the impact of fragmentation on the loss of specialist species, which are likely to be those of greatest conservation concern. By the same token, they may be overestimating the value of small fragments. At the extreme you could have a situation whereby following a fragmentation event, a set of newly created forest fragments undergoes complete turnover of species, replacing all the forest specialists with matrix generalist and edge species. The *z* value of the ISAR before and after turnover might remain relatively consistent, indicating no biodiversity loss through time and masking the loss of the forest specialists (Banks-Leite *et al.,* 2012). It is also possible that this 'masking effect' is impeding our ability to detect other ecological patterns. For instance, Krauss *et al.,* (2010) found evidence of an extinction debt in grassland vascular plant specialists, but not for non-specialised species, in a study of 147 European grasslands. However, it always has to be recalled that ISARs describe merely how many species are found on average in a patch or island of a particular size and do not describe

the degree of compositional overlap or beta diversity across the system. Should the distribution of species be highly nested then it follows that there should be a close correspondence between ISAR form and species loss rates, but if the distributions are strongly non-nested then there may be decoupling of extinction risk from ISAR form (see discussion in Whittaker and Fernández-Palacios, 2007).

In sum, a focus on total richness in ISAR studies can be misleading as the sensitivity of different species groups to habitat area shows high variability. A deconstruction approach (Marquet *et al.*, 2004), whereby the total species compliment is split into constituent subsets (e.g. based on trophic status or habitat specialisation), has been advocated as a useful method for ecological and conservation oriented research, but is regularly overlooked in ISAR studies (but see Bommarco *et al.*, 2010). Our results indicate that predictions of specialist species loss following fragmentation are too moderate in the absence of a deconstruction approach. This finding is especially concerning as these are the species that are most sensitive to fragmentation. Therefore, we argue it is essential to adopt a deconstruction approach if the ISAR is to be an effective conservation tool.

CHAPTER 8

Quantifying and interpreting nestedness in habitat islands: a synthetic analysis of multiple datasets

Paper in accepted in Diversity & Distributions

8.1 Preface

This chapter is the fifth empirical chapter, and examines the prevalence of nestedness in habitat islands. Most previous nestedness meta-analyses have included large numbers of oceanic island datasets, and so this represents the first analysis of nestedness in habitat islands using currently accepted methods. Nestedness was measured using NODF and the PP null model. Most datasets were not found to be significantly nested. The implications of these nestedness results on strategic conservation planning were evaluated using a minimum set reserve problem solving algorithm.

Authors: Thomas J. Matthews, H. Eden W. Cottee-Jones and Robert J. Whittaker

The co-author contributions were as follows. T.J.M. and R.J.W. conceived of and designed the study, T.J.M. collected and analyzed the data, with habitat specialization data contributed by H.E.C-J, and T.J.M. and R.J.W wrote the paper with significant input from H.E.C-J.

8.2 Abstract

Aim The concept of nestedness is important in determining the relative contribution to overall system diversity of different habitat patches within a fragmented system. Much of the previous work on nestedness has focused on islands within oceans (islands sensu stricto). The largest analysis of habitat island systems to date found significant nestedness to be a near universal feature, but the methods used have since been criticized as inappropriate. Thus, there is a need for an updated, critical examination of the prevalence, underlying drivers and implications of nestedness in multiple habitat island systems.

Location Global.

Methods Here we collate 97 datasets from published habitat island studies, comprising multiple taxa. We use the NODF metric (nestedness metric based on overlap and decreasing fill) to estimate nestedness, and determine significance using the four-step proportional-proportional algorithm to simulate presences/absence matrices. We investigate the role of habitat island area in driving observed nestedness. We use linear modelling to examine the impact of dataset characteristics on the degree of nestedness, and assess the conservational biogeographic implications of nestedness in relation to strategic conservation planning.

Results Significant nestedness occurred in only 9% of systems, while anti-nestedness (i.e. datasets less nested than expected by chance) occurred in 16% of systems. For the majority of datasets found to be significantly nested, we observed a relationship to fragment area, suggesting that structured extinctions may be important in determining the composition of certain habitat island communities. We found that the degree of nestedness in an archipelago is an important consideration for systematic conservation planning.

Main conclusions Significant nestedness is considerably less common in habitat islands than previously reported. Strategic guidance for conservation planning should proceed on a case

by case basis, and previous conservation recommendations based on the assumption of significant nestedness in most fragmented landscapes may need to be re-evaluated.

8.3 Introduction

The destruction and fragmentation of natural habitats is generally regarded as the largest driver of the current terrestrial extinction crisis (Sala *et al.*, 2000; Whittaker & Fernández-Palacios, 2007). The practice of deforestation and land-use change generally produces fragments of original habitat (i.e. habitat islands) set in a matrix of modified habitat. To understand and mitigate the impacts of this habitat loss on biodiversity many researchers have applied methods derived from the field of island biogeography (Diamond, 1975a; Whittaker & Fernández-Palacios, 2007; Ladle & Whittaker, 2011; Wang *et al.*, 2013).

The overall richness of a set of fragments in a disturbed landscape depends not only on the form of the island species–area relationship (ISAR; sensu Whittaker & Fernández-Palacios, 2007; Triantis *et al.*, 2012) but also on the compositional overlap between fragments. While the idea was discussed earlier (e.g. Darlington, 1957), nested structure was first formally defined in the 1980s as the situation in which depauperate island faunas constitute proper subsets of the species in richer islands (Patterson & Atmar, 1986). This definition was based on the idea of ordering a presence/absence matrix of species on islands in relation to island species richness. However, subsequent authors have used other variables (e.g. area) to order the presence/absence matrix (e.g. Kadmon, 1995; Wang *et al.*, 2010, 2013). Hence, nestedness may now be regarded simply as a form of ordered composition of species assemblages involving a significant tendency for packing of the matrix into a series of proper subsets. A variety of mechanisms has been identified as potentially contributing to patterns of nestedness or anti-nestedness differing from random expectation (see Table 8.1). Anti-nestedness is a term that encompasses several patterns of community structure: high turnover (Leibold & Mikkelson, 2002), perfect checkerboards (Diamond, 1975b), the

deviation from nestedness whereby within a set of sites species are only present in a single site (Poulin & Guégan, 2000), or simply to describe a dataset which is significantly less nested than expected by chance (e.g. Gotelli & Ulrich, 2012; Ulrich & Gotelli, 2012, 2013) (Table 8.1). This latter all-embracing definition is the definition of anti-nestedness used herein.

A number of nestedness metrics have been proposed (Atmar & Patterson, 1993; Wright *et al.,* 1998; Almeida-Neto *et al.,* 2008; Ulrich *et al.,* 2009), and whilst debate is ongoing, the nestedness metric based on overlap and decreasing fill (NODF) (Almeida-Neto *et al.,* 2008; see Table 8.2) is generally considered one of the most appropriate nestedness metrics (Almeida-Neto *et al.,* 2008; Ulrich & Almeida-Neto, 2012; Wang *et al.,* 2013). NODF is based on the twin properties of standardized differences in row and column fills and the overlap of presences in two adjacent columns (Table 8.2). There has also been debate about how to determine the significance of observed departures from random expectations. The usual protocol has been to compare the observed nestedness metric value with a distribution of values generated using a null model, with much debate arising about the choice of algorithms (Wright *et al.,* 1998; Miklós & Podani, 2004; Ulrich & Gotelli, 2007, 2012, 2013; Gotelli & Ulrich, 2012; Strona & Fattorini, 2014). Early studies found significant nestedness to be near universal (e.g. Simberloff & Martin, 1991; Atmar & Patterson, 1993) and it was thus generally assumed that nestedness was a common pattern in oceanic island systems (Wright *et al.,* 1998; Whittaker & Fernández-Palacios, 2007), and by extension, habitat islands. However, the majority of these early studies used null model algorithms based on the concept of random fill (Table 8.2), an approach which has since been criticized as inappropriate due to an inflation of type I errors (Ulrich & Gotelli, 2007; Gotelli & Ulrich, 2012). Random fill algorithms place no constraints on column and row totals (i.e. site richness and species incidence values), and thus marginal totals vary between simulated

matrices in no systematic manner. More appropriate null algorithms include those which keep row and column totals fixed ('fixed-fixed' algorithms; herein FF; Miklos & Podani, 2004), and an algorithm which varies row and column totals, but for which the average totals for a set of random matrices match the observed row and column totals of the empirical matrix ('proportional-proportional' algorithm; herein PP; Ulrich & Gotelli, 2012). The application of more appropriate algorithms to numerous datasets has led workers to question whether significant nestedness is as common as once assumed (e.g. Ulrich & Gotelli, 2007, 2012, 2013; Almeida-Neto *et al.*, 2008).

The choice between PP and FF depends on the type of data being analysed (Ulrich & Gotelli, 2012). The PP algorithm characterizes a biological situation in which island richness and species incidences vary between communities and through time (e.g. due to random extinction and colonization), but for which the average marginal totals match the empirical matrix. Random colonization and extinction through time may reasonably be assumed to be a characteristic of habitat island systems, which have more variable dynamics than more isolated continental shelf and oceanic islands (Ladle & Whittaker, 2011). Thus, for the present study we used the PP algorithm. Additionally, many habitat island datasets contain relatively small islands, and the PP model is preferred when the scale of analysis is small (Ulrich & Gotelli, 2012). Finally, metrics that use marginal totals for calculation, such as NODF, do not perform well with FF algorithms, which constrain totals (Ulrich & Gotelli, 2012, 2013).

Table 8.1 The different factors proposed to contribute towards nestedness or conversely to anti-nestedness in islands and habitat islands. * represents references that discuss the factor, but not in the context of nestedness or anti-nestedness.

Factors	Explanation/Mechanism	Examples and relevant references
Area	Nestedness is the result of differential area requirements of species. Species with smaller population sizes and larger area requirements have a greater risk of extinction and thus a predictable sequence of extinction occurs in relation to island size.	Patterson & Atmar (1986); Wright *et al.* (1998) [both for nestedness]
	Anti-nestedness related to area might hypothetically relate to assembly rules mediated through competitive interactions.	See below under Assembly rules
Isolation	Nestedness is due to predictable dispersal limitation, such that nestedness occurs due to differential immigration to islands.	Kadmon (1995)
	In contrast, diminishing nestedness can result from increased inter-patch distances that increase the likelihood of different islands sampling different species pools, and of islands being occupied by different subsets of the regional species pool (or pools) due to differences in, for example, climate and geology.	
Habitat nestedness	Habitats exhibit a nested pattern and there is a strong affinity between species and habitats, generating nestedness.	Honnay *et al.* (1999) Triantis & Bhagwat (2011)
Habitat quality	Nestedness is the result of differences in species' tolerance to habitat quality combined with variation in the quality of habitat amongst patches of the same habitat.	
Passive sampling	Nestedness structure can simply be the result of a sampling effect: if islands draw species from the regional pool and the pool follows a particular species abundance distribution (e.g. lognormal).	Cutler (1994); Whittaker & Fernández-Palacios (2007)
In situ speciation	Speciation occurring within individual islands will result in single island endemic species and thus will diminish the likelihood of nestedness and increase the likelihood of significant anti-nestedness in a set of islands. This is more likely to occur in oceanic islands than in habitat islands.	Whittaker & Fernández-Palacios (2007)*
Different species pools	If different islands within an archipelago draw species from different source pools, the degree of nestedness in the archipelago will be reduced.	Whittaker & Fernández-Palacios (2007)*
Human factors	Activities such as hunting may result in an ordered loss of particular species and thus increase nestedness in an	Ladle & Whittaker (2011)*

	archipelago, whereas species introductions may be island specific and thus decrease nestedness.	
Disturbance	Differences in disturbance regimes across an archipelago can increase or decrease nestedness.	Fleishman & Murphy (1999); Wang *et al.* (2013)
Assembly rules	Diamond (1975b) hypothesized (controversially, as it turned out) that, when focusing on particular guilds, the occurrence of particular assembly rules in an archipelago can result in ordered patterns of community structure; for example, competitive interactions may generate differences in assemblage composition from island to island in a fashion not explicable simply by island area. In this vein, the occurrence of supertramps species (i.e. species that have a higher incidence on species poor islands) will potentially reduce nestedness in an archipelago.	Diamond (1975b)

Previous meta-analyses of insular nestedness (e.g. Wright *et al.*, 1998; Ulrich & Gotelli, 2007, 2013; Almeida-Neto *et al.*, 2008) have relied largely upon Atmar & Patterson's (1995) collection of presences/absence matrices. This collection contains a large number of oceanic and continental shelf island datasets, and not so many non-experimental habitat island systems. The only meta-analysis of nestedness patterns which included a relatively large number of habitat island datasets to date, by Watling & Donnelly (2006), found near universal (94%) significant nestedness in the datasets examined; although this study still included some oceanic island datasets. However, these authors used a metric (nestedness temperature) and a random fill algorithm, both of which have since been criticized as being prone to type I errors, i.e. a tendency to greatly over-estimate the degree of nestedness (above). Thus, there is an exigent need to apply more appropriate methods to determine the prevalence of nestedness/anti-nestedness in a larger number of habitat island systems (cf. Sekercioglu & Sodhi, 2007).

The prevalence of nestedness (and of anti-nestedness) in habitat islands is important from both a fundamental ecological and conservation biogeographic perspective because it potentially informs protected area placement and design in fragmented and degraded

landscapes (Cutler, 1994; Fischer & Lindenmayer, 2005a, b; Triantis & Bhagwat, 2011), contributing to the SLOSS ('single large or several small') protected areas debate (see Ovaskainen, 2002), and the minimum set problem (herein, MS). The MS problem is a key component of strategic conservation planning (Watson *et al.*, 2011), and describes the search for a solution to the problem of what is the smallest number of sites/protected areas within a region needed to ensure each species in the region is represented at least once (see Watson *et al.*, 2011). Nestedness is thus linked to the MS solution. For example, if a set of habitat islands in a dataset is perfectly nested according to island area, all the species in the islands can be represented at least once by conserving the largest habitat island. However, as perfect nestedness is extremely uncommon the situation is rarely so simple (Fischer & Lindenmayer, 2005b). The link between nestedness and MS has generally been examined only within individual habitat island datasets (e.g. Fischer & Lindenmayer, 2005a), and a synthetic analysis of a large number of empirical habitat island datasets is warranted.

There is also a need to improve understanding of the mechanisms driving nestedness where it does occur in habitat island systems (Fischer & Lindenmayer, 2005b; Morrison, 2013) as, from a conservation perspective, the identification of pattern is of limited utility without an understanding of underlying process. Habitat island area has long been known to be one of the most important variables underpinning species composition patterns in fragmented landscapes (Watling & Donnelly, 2006; Whittaker & Fernández-Palacios, 2007), and thus a rigorous analysis of the role of island area in driving nestedness patterns in habitat islands is needed.

Finally, there is a need to assess whether nestedness patterns in species of conservation interest (i.e. habitat specialists) are masked by the inclusion of generalist species. For example, habitat generalists have been found to depress the slope of the ISAR for birds in forest fragments (Matthews *et al.*, 2014a). This type of deconstruction approach,

whereby the total species compliment is split into constituent subsets (e.g. based on habitat specialization), has been advocated as a useful method for conservation oriented research, but has only rarely been applied in nestedness studies (e.g. Blake, 1991; Fischer & Lindenmayer, 2005a).

Here, we undertake a synthetic analysis of 97 habitat island datasets with four aims. First, we quantify the prevalence of nestedness and anti-nestedness in habitat island systems (including several datasets not incorporated in past meta-analyses) using statistically appropriate methods. Second, we investigate the role of island area (Table 8.1) in observed nestedness patterns. Third, we examine whether habitat specialists and generalists exhibit different nestedness patterns. Finally, we investigate the conservation biogeographic implications of nestedness in our datasets by exploring the linkages between nestedness and the MS problem.

8.4 Methods

8.4.1 Data collection

We searched for relevant habitat island studies and datasets using several abstracting databases between May 2010 and August 2013. Search keywords included combinations of 'habitat islands', 'species richness', and 'fragments'. Certain datasets were obtained from the authors of the source papers, whilst others were supplemented with additional data from the source paper authors. Following Matthews *et al.* (2014b), datasets were included based on the following criteria:

1. habitat islands were defined as discrete patches of habitat surrounded by contrasting matrix habitat(we also included a small number of datasets consisting of protected areas in which the contrast between the islands and the intervening matrix was not so pronounced);
2. there were at least six habitat islands;
3. the area and species richness of each habitat island were given; and

4. data from any particular study did not overlap with those from any other study accepted for analysis (data for different taxa within the same study system were allowed).

For the purposes of this study, we considered a few island systems within a water matrix (e.g. flooded rainforest fragment systems) as habitat islands, as the range of island areas and dominant assembly processes are more similar to habitat islands than oceanic islands. For each dataset we recorded a variety of system characteristics: the habitat island type (i.e. forested or non-forested), the taxon studied, the latitude of the centre of the study extent, the range of species richness, and the habitat island sizes (Table A6.1 in Appendix 6). A small number of datasets included sites which contained zero species. We removed these sites prior to analysis.

8.4.2 Quantifying nestedness

To quantify nestedness in our datasets we used the NODF metric as it is widely regarded as the most statistically appropriate (Almeida-Neto *et al.*, 2008; Morrison, 2013; see Table 8.2). NODF was calculated using the 'vegan' R package (Oksanen *et al.*, 2013). The NODF metric allows nestedness to be calculated independently for matrix rows (i.e. nestedness amongst sites) and matrix columns (i.e. nestedness amongst species incidences), as well as combined for the whole matrix (i.e. maximally packed nestedness; herein 'max matrix'). Following Morrison (2013), we calculated all three NODF values and compared the value for rows (sites in our matrices when using vegan) with that for columns for each dataset, taking the larger of the two values to indicate that a particular type of nestedness contributed more to the overall nestedness pattern.

Table 8.2 The nestedness metric employed in the present study, along with the two null community simulation algorithms used to determine whether the degree of nestedness was significant for a given system.

Metric	Abbreviation	Description	References
Nestedness metric based on overlap and decreasing fill	NODF	Based on the twin properties of standardized differences in row and column fills and paired overlap (i.e. the overlap of presences in two adjacent columns). Can quantify nestedness for the whole matrix and for rows and columns separately. Increasing NODF equates to increasing nestedness.	Almeida-Neto *et al.* (2008)
Simulation algorithm		Description	References
Proportional-Proportional (PP)	PP	When a set of null matrices are simulated using the PP algorithm, the column and row totals do not all match the totals of the empirical matrix. Rather, the average totals for the set match the totals of the empirical matrix. The PP algorithm is more ecologically realistic in the context of habitat islands and small patches as stochastic temporal variation in island richness and species incidences (i.e. the biological equivalent of variations in marginal totals) is a frequently observed characteristic of such systems. Both matrix fill and the matrix dimensions are kept constant.	Ulrich & Gotelli (2012)
R00	R00	Keeps the number of presences constant, but allows row and columns totals to vary in an equiprobable manner. This algorithm has been criticized for high prevalence of type I errors, and we use it in the present study simply to enable comparisons with previous studies of nestedness in habitat islands.	Wright *et al.* (1998)

8.4.3 Determining the significance of nestedness patterns

To determine if the observed NODF values were significantly different from values expected for a randomly assembled community, we simulated 1000 binary presence/absence communities in each case, recording the metric values for each simulation. To simulate the null model communities we used the aforementioned four-step PP algorithm (Ulrich & Gotelli, 2012; see Table 8.2). To enable a rough comparison between our results and those of Watling & Donnelly (2006), we re-ran our analyses using the R00 algorithm (Table 8.2). We used R00 simply because Watling & Donnelly used a random fill algorithm, and unless R00

is specifically mentioned, significant relationships reported below were determined using the PP algorithm. The PP communities were simulated using the 'NODF program' (Almeida-Neto & Ulrich, 2010), and we used the one-sided *P* value generated from the NODF program to determine significance (i.e. *P* <

0.05), which is based on the assumption of a normally distributed Z-transformed score (discussed below). It should be noted that in contrast to our 'vegan' NODF analyses, the NODF program analyses matrices in which species are rows. The R00 communities were simulated using the 'vegan' R package, and we used the default *P* value generated by vegan.

8.4.4 Does island area underpin observed nestedness in habitat islands?

For each dataset found to be significantly nested according to PP, we first calculated the maximum NODF value (row orders were not kept constant). We then ordered the matrices by decreasing area (i.e. largest island as the top row) and calculated NODF after holding the row orders constant. This enabled us to compare the row orders of the matrices ordered by area with the row orders of the max matrix to determine if the two were significantly correlated using the standard Spearman's correlation test, and thus whether area may be driving any observed nestedness pattern (cf. Schouten *et al.,* 2007; Wang *et al.,* 2010). In addition to analysing area-ordered nestedness, we originally planned to investigate whether island isolation, habitat nestedness and passive sampling (Table 8.1) were also driving any observed nestedness in four datasets sourced from our previous work (Matthews *et al.,* 2014a). However, these datasets were not found to be significantly nested according to the PP algorithm (discussed below) and we were thus unable to proceed with this idea. Nonetheless, in preparation for this work we developed and coded a model based on Coleman's (1981) passive sampling model in R as it was not available elsewhere. To aid the community, we present the code alongside example passive sampling plots in Appendix 6.

8.4.5 Habitat specialization

For a separate study (Matthews *et al.,* 2014a), we took 16 of the 97 datasets pertaining to birds in forest fragments within an agricultural matrix and classified all bird species (over 1000 in total) as forest generalists or specialists using an extensive classification methodology (see Matthews *et al.*, 2014a). In the present paper, we used that classification and calculated the NODF *Z*-transformed score (hereafter simply *Z*-score) for generalists and specialists separately for the 17 datasets and compared the *Z*-scores, where $Z = \frac{\mathrm{Obs} - \mathrm{mu}}{\mathrm{SD}}$, and where Obs is the observed nestedness value according to a given metric, mu is the mean nestedness metric value of the PP simulated communities (based on 1,000 simulations), and SD is the standard deviation of the 1,000 values. The *Z*-score was used in preference to the observed NODF value as the former represents a standardized effect size and can thus be used in comparative analysis (e.g. Ulrich & Gotelli, 2013).

8.4.6 Model selection and dataset characteristics

We fitted generalised linear models (GLMs) with Gaussian variance, using six dataset characteristics as the predictor variables (number of islands, number of species, minimum island area, taxon studied, habitat island type, latitude). We took the absolute value of latitude. For the response variable we used the *Z*-score. The assumptions of linear modelling (i.e. predictor normality, no outliers, and minimal multicollinearity) were all tested, and number of species, number of islands and the minimum island area were each log-transformed (natural logarithms were used) prior to analysis. One dataset was removed as an outlier. The model with the lowest Akaike's Information Criterion corrected for sample size (AIC_c; Burnham & Anderson, 2002) was considered to be the best model. However, if any model was within Δ AIC_c of < 2 of the best model, we considered that model as having a similar degree of support (Burnham & Anderson, 2002). We determined the weight of

evidence for each variable by summing the Akaike weights of all the models in which a variable was included (cf. Brook *et al.,* 2006). We fitted a complete set of models, considering all predictors, using the 'dredge' function in the 'MuMIn' R package (Bartoń, 2012). Our GLMs revealed that the number of species in a dataset was a correlate (potentially a driver) of the *Z*-score (discussed below). Bivariate plotting indicated a possible boundary effect in the top right quadrant, and thus we conducted boundary tests using ECOSIM (Gotelli & Entsminger, 2001b).

8.4.7 Minimum set problem

To examine the conservation implications of nestedness in terms of protected area network design (Triantis & Bhagwat, 2011), we developed an algorithm to solve a MS reserve location problem (Watson *et al.,* 2011). That is, for each dataset we determined the smallest number of habitat islands required in order for all species in the dataset to be represented at least once. For a given dataset the algorithm worked by first selecting the islands that included singleton species (i.e. species only present on one island in the dataset), storing the island and species identities, and then removing these islands and species. From the remaining islands, the most species rich island (if there were ties, the first site was taken) was then selected and the unique species names (i.e. species not stored in previous iterations) present on this island were stored. This island was then removed, and the process repeated with the next most species-rich island, and so on until all species in the dataset had been recorded at least once. This algorithm was then applied to each dataset individually. Finally, we adapted the MS algorithm to determine what proportion of a dataset's species was represented in just the largest island. Unless otherwise stated, analyses were conducted in R (R Development Core Team, 2013).

8.5 Results

Over 1,000 published articles were screened, of which 97 were deemed suitable for analysis (Fig. 8.1, Table A6.1 in Appendix 6). These 97 datasets comprised 69 vertebrate, 20 invertebrate, and 8 plant datasets; and 70 forested habitat islands, and 27 non-forested habitat islands.

8.5.1 Prevalence of nestedness in habitat islands

Considering all 97 habitat island datasets, the NODF values tended towards the nested end of the NODF spectrum, that is the values were generally closer to 100 than 0 (mean value=64.6; range=36–88). However, 62 (64%) datasets had negative *Z*-scores (Table A6.2 in Appendix 6). Based on the PP algorithm, it appears that significant nestedness and anti-nestedness are relatively uncommon in habitat islands. Nine datasets (9%) were found to be significantly nested, and 16 datasets (16%) were significantly anti-nested (an example of each is provided in Fig. 8.2); the remaining 72 datasets (74%) had NODF values not significantly different from random. For eight of the nine significantly nested datasets the NODF value was higher when calculated for matrix rows (i.e. by sites) than for matrix columns (Table A6.3 in Appendix 6). Considering all datasets, the NODF value was higher when calculated for matrix rows for 83 (86%) datasets (Table A6.3).

When R00 was used to simulate presence/absence matrices, significant nestedness was almost a universal finding, i.e. 96 datasets were deemed to be significantly nested (Table A6.2).

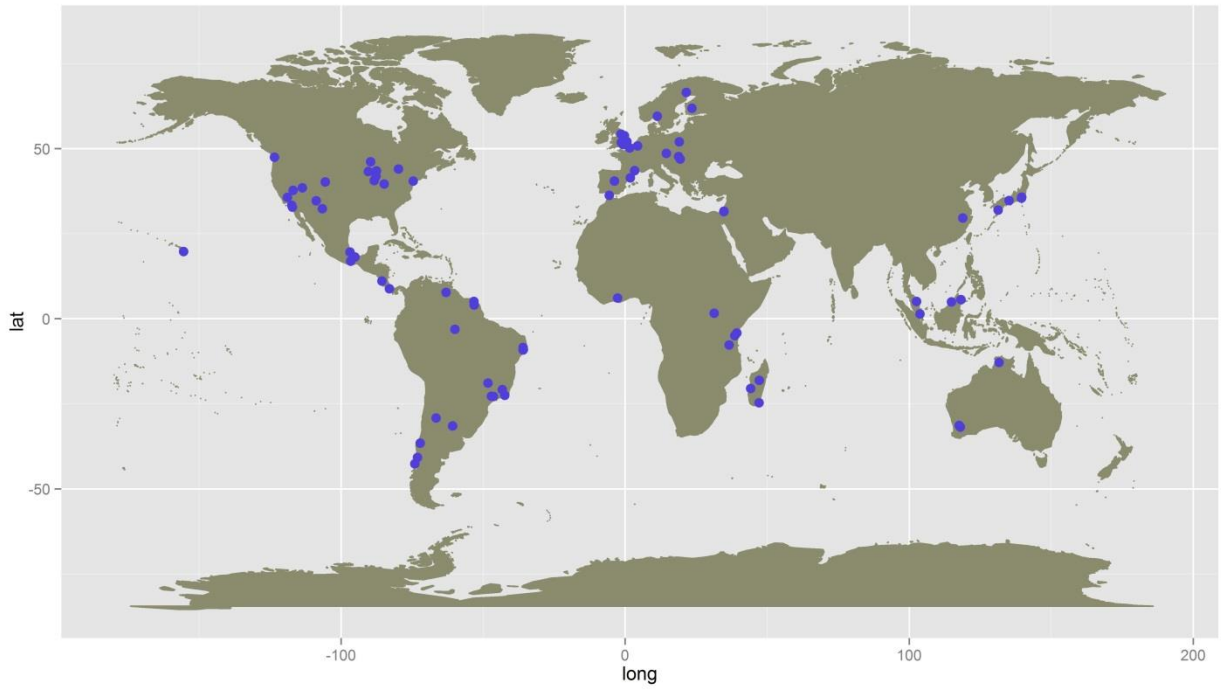


Figure 8.1 A map of the 97 datasets (blue dots) included in this study. Multiple datasets based in the same location are represented by a single blue dot. Where possible we took the coordinates from the centre of the study extent. However, in a small number of papers only a general area was listed (e.g. East Central Illinois), and in these instances we simply used the centre of this area.

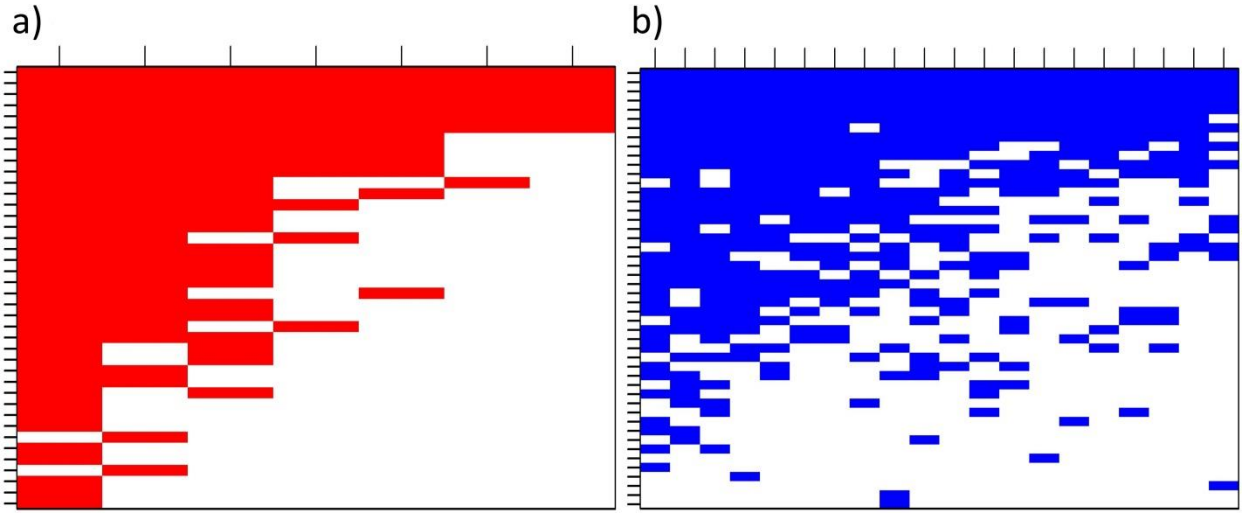


Figure 8.2 Incidence plots for two habitat island datasets: a significantly nested dataset (a), and a significantly anti-nested dataset (b). The nested metric employed was NODF (Almeida-Neto *et al.*, 2008) using the maximally packed matrix. The x-axis represents the individual habitat islands, and the y-axis represents individual species. The coloured bars thus indicate that a given species was sampled in a given site, while a blank bar indicates that a species was not found in a site. Significance was determined in each case by comparing the observed NODF statistic with a distribution of values generated by 1000 null communities (using the PP null model). (a) A dataset of frogs in forest fragments, Brazil (number of species = 40; number of islands = 7; NODF value = 80; *Z*-score =1.74; Zimmerman & Bierregaard, 1986), and (b) is a dataset of birds in forest fragments, UK (number of species = 48; number of islands = 20; NODF value = 70; *Z*-score =-2.11; Ford, 1987).

8.5.2 Mechanisms and dataset characteristics

When considering all 97 datasets, the row order of the area-ordered matrix was significantly correlated with that of the max matrix for 53 datasets (55%; Table A6.4 in Appendix 6). When considering the nine significantly nested datasets, the row order of the area-ordered matrix was significantly correlated with that of the max matrix for eight datasets (89%; mean Spearman's Rho = 0.87; mean Rho for all nine datasets = 0.81; Table 8.3).

Habitat specialization

Within the analysis of the 16 datasets for which bird species were divided into habitat specialists and generalists, specialists had a greater *Z*-score for 11 datasets (69%), but specialist *Z*-scores were negative in 6 cases, indicating anti-nestedness in a number of datasets (Table A6.5 in Appendix 6). Moreover, the difference between generalist and specialist subset *Z*-scores was marginally significant according Wilcoxon rank test (W = 76, *P*=0.05).

Table 8.3 The correlation of the row orders of 9 habitat island presence/absence matrices ordered according to decreasing area (Area NODF), with the row orders of the maximally packed NODF matrices (Max NODF). Correlation was determined using Spearman's correlation and the correlation coefficient is presented with the *P* values. These nine datasets are those that were found to be significantly nested, out of the 97 habitat island datasets examined. The max NODF value and Area NODF value given in the table are the NODF value for matrix rows (i.e. nestedness amongst sites). The significance of the observed NODF value was determined by comparing the observed value with a distribution of values obtained for 1000 simulated null communities, using the PP null model. The correlation results for all 97 datasets are presented in Table A6.4 in Appendix 6. The dataset numbers correspond to the dataset information in Table A6.1 in Appendix 6.

Dataset	Max NODF	Area NODF	Spearman's Rho	*P*
4.	91.55	90.03	0.99	<0.01
25.	84.18	79.71	0.87	<0.01
30.	85.12	79.23	0.95	<0.01
41.	73.84	64.81	0.38	0.16
52.	87.81	85.43	0.98	<0.01
75.	89.68	88.01	0.99	<0.01
83	80.60	72.89	0.91	<0.01
84.	75.26	58.52	0.40	0.04
97.	95.82	82.90	0.89	0.01

Model selection results

When NODF was used to calculate the *Z*-score across all datasets, the best model contained only the number of species (Table 8.4). Number of species had a relatively high weight of evidence value. As an increasing *Z*-score implies increasing nestedness according to NODF, this result indicates that datasets with more species are less nested. However, the best model had an adjusted R^2 value of only 0.20. Whilst a bivariate plot indicated a possible boundary effect in the upper right quadrant, there were no fewer data points in the upper right quadrant than expected by chance according to a boundary test when the sum of squares criterion was used (P = 0.17; see Fig. A6.1 in Appendix 6), although when the number of points criterion was used the result was marginally significant (P =0.05). Latitude and island type were included in models within <2 ΔAIC_c of the best model, but had relatively low weight of

evidence values. Taxon was never in the best models, and had a low weight of evidence value (Table 8.4).

Table 8.4 Parameter estimates of a set of parsimonious generalised linear models with Gaussian variance, for 97 habitat island datasets. The best model (i.e. lowest AIC_c) and all models within approximately ΔAIC_c of < 2 of the best model are given. The predictor variables included the area of the smallest island (Min. Area), the number of islands (No. Frag.), the number of species (No. Sp.), the taxon (Taxon), the habitat island type (i.e. forest or non-forest; Hab. Type), and latitude of the study area (Lat.). The response variable was the *Z*-score (calculated using NODF) for each dataset, calculated by simulating 1000 null community matrices using the PP algorithm. The number of model parameters (Par.), the delta AIC_c (ΔAIC_c), and the Akaike weights ($wAIC_c$) are also presented. Following Brook *et al.* (2006) the weight of evidence of each variable, calculated by summing the Akaike weights of all the models in which a variable was included, is also given. NI (not included) indicates a variable was not included in a model.

Model Rank	Hab. Type	Min. Area	No. Frag.	No. Sp.	Taxon	Lat.	ΔAIC_c	$wAIC_c$	Par.
1	NI	NI	NI	-0.62	NI	NI	0	0.20	1
2	NI	NI	0.17	-0.63	NI	NI	1.27	0.11	2
3	+	NI	NI	-0.61	NI	NI	1.46	0.10	2
4	NI	NI	NI	-0.63	NI	<-0.01	1.85	0.08	2
Weight of evidence	0.33	0.25	0.37	1.00	0.09	0.32			

8.5.3 Minimum set results

The mean proportion (Prop) of islands required to represent each species at least once (the minimum set: MS) was 40.8% (results for individual datasets are presented in Table A6.2). Six datasets (6%) required only a single island; for four of these cases, the island in question was the largest island. In contrast, six datasets (6%) needed all islands to represent every species. As expected, the MS results had strong concordance with the nestedness results. The NODF value was significantly negatively correlated with Prop (-0.45, P = <0.001; Fig. 8.3a). All datasets which required more than 40% of sites to represent all species (N=34; 35%) had a negative *Z*-score, indicating anti-nestedness. The mean Prop for significantly anti-nested datasets was 53.1%, whilst for significantly nested datasets the mean Prop was only 17.7%.

The mean proportion of species represented when only the largest patch in a dataset was considered was 63.8% (range = 21.0 to 100%; Table A6.2).

A Kruskal–Wallis test indicated that Prop significantly differed between taxa (16.2, *P* = < 0.01), and pairwise Wilcoxon tests indicated that this difference was driven by vertebrates (Fig. 8.3b). Vertebrates had a significantly lower Prop than both plants (*W*=456.5, *P* = <0.01), and invertebrates (*W*=377.5, *P* = <0.01). Prop did not significantly differ between invertebrates and plants (*W*=98.5, *P* =0.36)

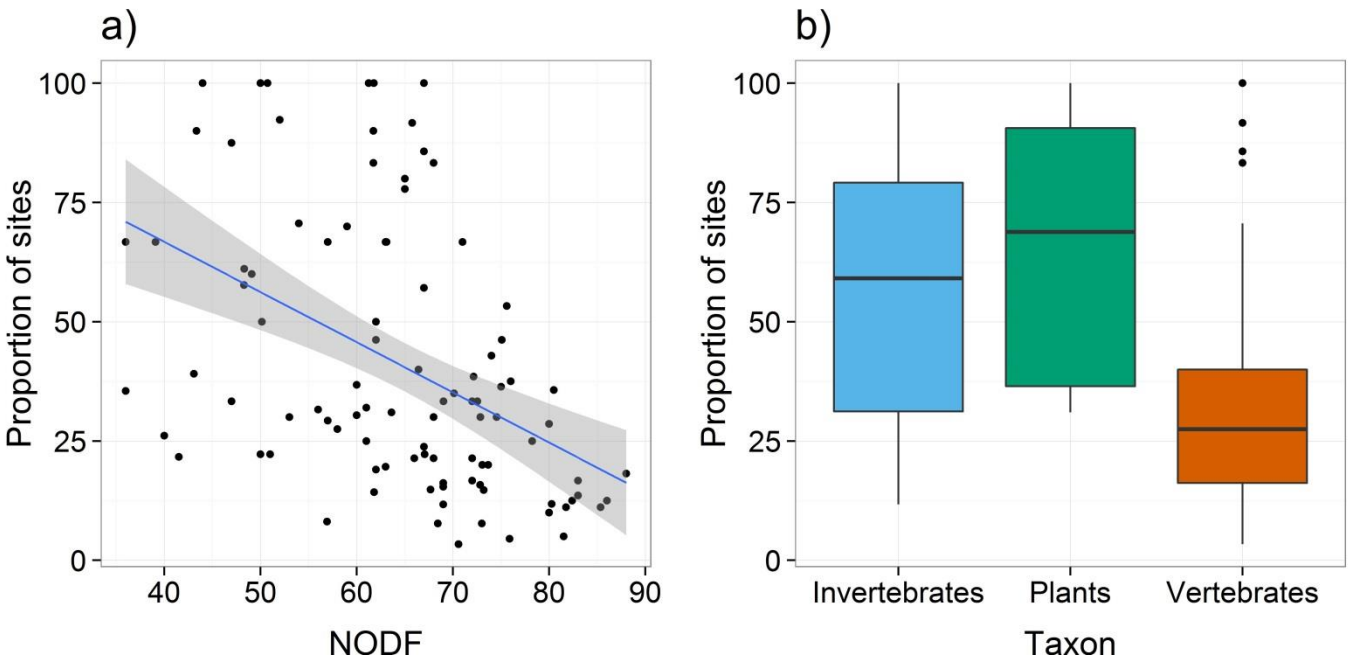


Figure 8.3 The relationship between (a) the smallest proportion of sites required to represent all species in a dataset (i.e. the solution to the minimum set problem), and the NODF value (maximally packed matrix), and (b) the variation in this proportion across taxa, for 97 habitat island datasets. To determine the solution to the minimum set problem we first ran an algorithm to determine the smallest number of habitat islands required in order to include all the species in a dataset. This number was then represented as a proportion of the total number of sites in the dataset. In (a), the blue line represents the fit of a standard linear model, and the grey shading represents the 95% confidence interval around this line. The box plots display the median (thick black line), the first and third quartiles (thin black box). The whiskers extend from the hinge to the highest value that is within 1.5 multiplied by the inter quartile range, of the hinge. Outliers are indicate by black circles.

8.6 Discussion

We have undertaken the first synthetic analysis of nestedness patterns in a large number of habitat island datasets using currently recommended statistical methods. Quantifying nestedness in habitat islands is important as it provides information on how species are distributed in patchy landscapes: information that can potentially be used to develop conservation strategies (Triantis & Bhagwat, 2011). Using the four-step PP null model algorithm (Ulrich & Gotelli, 2012) we found that significant nestedness was only apparent in 9% of datasets, whilst 16% of datasets were significantly anti-nested. Notwithstanding the fact that most datasets were not significantly nested, we have also demonstrated that absolute habitat island area appears to be an important environmental factor correlating with observed nestedness patterns.

8.6.1 Nestedness prevalence and mechanisms

Our finding that nestedness is very much the exception rather than the rule in habitat island datasets contradicts a previous analysis incorporating multiple habitat island datasets (Watling & Donnelly, 2006) but is consistent with recent studies that have incorporated habitat island datasets alongside a larger number of oceanic, continental-shelf, and simulated island datasets (e.g. Ulrich & Gotelli, 2007; Ulrich & Gotelli, 2013). The core reason for the difference in our findings, which are based on NODF and the PP algorithm, and those of Watling & Donnelly, is their use of the nestedness temperature calculator which uses a random fill algorithm that has since been shown to be extremely prone to type I errors (see also Gotelli & Ulrich, 2007, 2012). This is clearly shown by the near universal significant nestedness indicated when we ran the R00 algorithm, for purely comparative purposes, on our 97 datasets (see Results and Table A6.2). Additionally, Watling & Donnelly used the nestedness temperature index, although re-running our analyses using nestedness temperature generates very similar results. Given that we found significant nestedness in only 9% of our

datasets using the preferred PP algorithm it follows that conservation managers should not assume significant nestedness *a priori* in fragmented systems.

The importance of island area in driving observed nestedness

Understanding the mechanisms and variables underpinning nested structure is arguably more important than simply quantifying nestedness, but these have been much less well studied, especially in habitat islands, as the relevant information cannot simply be gleaned from the observed nestedness values (Morrison, 2013). Whether considering either the significantly nested datasets, or all datasets, area seems to be an important environmental factor correlating with the observed extent of nestedness in our datasets. The potential significance of habitat island area was further indicated by the finding that the observed row-ordered NODF value was generally larger than the column-ordered value. This signifies that nestedness between sites is more important to the overall nestedness pattern than nestedness among species (Morrison, 2013). That being said, another potential explanation for larger row-ordered NODF values compared to column-ordered values relates to the fact that datasets generally contain a small number of singleton species (i.e. species present in only one site). By definition, there is no nestedness among singleton species. In contrast, very few datasets contain sites with only a single species. Thus, this finding may be due to the differences in the level of ties between species incidences (i.e. column totals) and the level of ties between species richness values (i.e. row totals).

Explanations for anti-nestedness (i.e. datasets less nested than expected by chance) require further exploration, reflecting that this phenomenon can reflect diverse patterns and causes (see Table 8.1; Ulrich & Gotelli, 2013). We did not examine the effect of area on significantly anti-nested datasets as a separate group (although the results are presented in our analysis of an area effect in all datasets; Table A6.4), for this very reason, i.e. because of the nebulous nature of anti-nestedness we were unsure what such an analysis would achieve.

Furthermore, multiple mechanisms, many of which require considerable data to examine (Table 8.1), may well be acting in tandem, with the effect of diluting any nestedness patterns (Ulrich & Gotelli, 2013). This dilution may partly explain the high number of datasets with non-significant NODF values. Additional possibilities underpinning the non-significant results are a) the varying responses of different types of species, namely generalists and specialists, to landscape composition, and b) the small number of islands in many datasets. On the whole, generalists were found to be less nested than specialists (discussed below) and thus datasets with a high proportion of generalist species may tend towards anti-nestedness. Other ecological types of species, especially transient species and those species that benefit from fragmentation (Blake, 1991; Fischer & Lindenmayer, 2005a) may also contribute towards anti-nestedness patterns.

The relationship between the Z-score and the number of species

The number of species was included in all GLMs with an $\Delta AIC_c < 2$, and the variable had a high weight of evidence value. However, further analysis revealed that the R^2 values of these models were low (roughly 0.20) and a boundary test revealed that there were not fewer data points in the upper right quadrant than expected by chance. Furthermore, in previous studies based on extensive simulations, the *Z*-score derived using both a different type of proportional null model and fixed-fixed null model has been found to be independent of matrix size (Almeida-Neto *et al.*, 2008; Strona & Fattorini, 2014). It thus appears unlikely that the *Z*-scores calculated in the present study are particularly biased in regards to the number of matrix columns.

8.6.2 Are habitat specialists more nested than generalists?

Several studies have found and argued that, as habitat specialists are generally more affected than generalists by habitat loss, specialists should exhibit a more nested structure (e.g. Blake, 1991; Fischer & Lindenmayer, 2005a; Schouten *et al.*, 2007; Sekercioglu & Sodhi, 2007).

For example, Blake (1991) observed that in Illinois woodlots forest core bird species were more nested than edge bird species. Our results are consistent with this assertion, although not entirely conclusively. Whilst specialists were more nested than generalists in 69% of cases, the specialists' subsets had a negative *Z*-score in 38% of datasets, indicating a tendency towards anti-nestedness. Based on these findings, further analyses of subsets of species may prove enlightening. Deconstructing species into subsets prior to data analysis has revealed interesting patterns of conservation relevance in ISARs (Matthews *et al.,* 2014a), and there is a need for a more comprehensive analysis of the impacts of, for example, generalist, migrant and transient species on nestedness patterns, particularly in fragmented landscapes.

8.6.3 Conservation implications

The mechanisms and environmental factors responsible for non-random assembly/disassembly patterns of (some) habitat island biotas remain poorly understood. According to our data, island area is an important environmental factor. This finding in isolation suggests that protecting few large habitat islands in a given region should be effective in conserving the majority of species in a region. However, when using the PP algorithm significant nestedness was uncommon, the area-ordered matrix was not always significantly correlated with the max matrix (e.g. Table 8.3), and moreover, we never observed perfect nestedness according to any metric (see also Fischer & Lindenmayer, 2005b). This point is further evidenced by the fact that our minimum set (MS) analyses showed that generally a high proportion of sites (mean = 40.8%) was required to represent all species in a dataset at least once. Intuitively one would expect the MS problem to be linked in some way with nestedness (Fischer & Lindenmayer, 2005b; Triantis & Bhagwat, 2011), and this is indeed the case in our habitat island datasets.

The goal of seeking representation of all elements (species, ecosystems, etc.) is a fundamental principle in strategic conservation planning (Watson *et al.,* 2011) and the link

between nestedness and the MS solution has been argued elsewhere, particularly in the context of the SLOSS debate (e.g. Whittaker & Fernández-Palacios, 2007). However, to our knowledge, it has not been empirically tested using a large number of habitat island datasets. In the context of SLOSS, the 'single-large' argument is based on the premise that a single large reserve is more effective than several small reserves of equivalent total area (Ovaskainen, 2002). Under a scenario of perfect, or at least high, nestedness-by-area, the largest island should include all of the species in the landscape. It is rarely so simple in practice. For example, in a study of forest birds in a fragmented landscape in Australia, Fischer & Lindenmayer (2005b) reported a pattern of habitat islands being significantly nested by island area. However, the islands were not perfectly nested and over a quarter of the study area was necessary in order to represent 80% of species classified as forest sensitive (Fischer & Lindenmayer, 2005b). In our analyses, the largest island contained all species in only four datasets, while on average the largest patch contained only 63.8% of the species in a dataset. These findings highlight that, in almost all instances, multiple islands of differing size are required in order to represent all species. In sum, we are unable to recommend a one-size-fits-all guideline for protected area design based on empirical patterns of nestedness.

The low number of plant (n=8) and invertebrate (n=20) datasets relative to vertebrate datasets (n=69) makes it problematic to discern a simple explanation for the difference in Prop (i.e. the mean proportion of islands required to represent each species at least once) between taxa (Fig. 8.3b), and further research is needed as this observation may have important conservation implications. It is important to note that our MS analyses are necessarily a simplification of the ecological reality as we only have static 'snap shots' of species composition. Whilst representation is a key component of strategic conservation planning, persistence is also important; the presence of a species in a site does not necessarily mean the species will survive at the site in the long run (Whittaker & Fernández-Palacios,

2007). Furthermore, our analyses assume that the habitat islands included in a dataset are the only islands in the landscape available for conservation: an unlikely situation in most cases. However, despite these simplifications the analyses provide a useful first approximation of the issues involved in assuming a specific protected area plan *a priori* (see also Fischer & Lindenmayer, 2005b).

Most nestedness meta-analyses have incorporated large numbers of oceanic and continental shelf island datasets. As habitat islands differ from such 'true' islands in many ways (Whittaker & Fernández-Palacios, 2007) it is important to look for general patterns within habitat island datasets when attempting to formulate conservation guidance. Using NODF and the PP algorithm, we found that only 9% of datasets were significantly nested. Hence, previous conservation recommendations based on the assumption of significant nestedness in most fragmented landscapes may need to be re-evaluated. Thus, we conclude that strategic guidance for conservation planning in fragmented landscapes should proceed on a case by case basis, rather than presuming any degree of nested structure *a priori*.

CHAPTER 9

Assessing functional nestedness patterns in habitat islands: an analysis of multiple bird datasets

Paper in review in Oikos

9.1 Preface

This chapter is the sixth empirical chapter, and tests the role of habitat island area in driving the functional nestedness of birds in habitat islands. Functional nestedness was measured using the recently proposed treeNODF metric, which takes into account how similar species are according to a functional dendrogram. Presence-absence matrices were ordered by island area to test for the effect of island area on any observed functional nestedness. Two null models were used to test for significance of the treeNODF values. The results were sensitive to the null model used. When a permutation model was used, significant functional nestedness was relatively common. However, when the more conservative fixed-fixed model was used, significant nestedness was much less prevalent.

Authors: Thomas J. Matthews, Catherine Sheard, H. Eden W. Cottee-Jones, Tom. P Bregman, Joseph A. Tobias and Robert J. Whittaker

The co-author contributions were as follows. T.J.M. and R.J.W. conceived of and designed the study, C.S., T.P.B., T.J.M. and H.E.C-J collected the data, T.J.M analyzed the data, and T.J.M. wrote the paper.

9.2 Abstract

The widespread destruction and fragmentation of natural habitats around the world creates a strong incentive to understand how species and communities respond to such pressures. The vast majority of research into habitat fragmentation has focused solely on species presence or absence. However, analyses using innovative functional methodologies offer the prospect of providing new insights into the key questions surrounding community structure in fragmented systems. A key topic in fragmentation research is nestedness (i.e. the ordered composition of species assemblages involving a significant tendency for packing of the presence–absence matrix into a series of proper subsets). To date, nestedness analyses have been concerned solely with nestedness of species membership. Here, we capitalize on the publication of a recent nestedness index (traitNODF) in which the branch lengths of functional dendrograms are incorporated into the standard NODF nestedness index. Using bird community data from 18 forest-habitat-island studies, and measurements of eight continuous functional traits from over 1000 bird species, we conduct the first synthetic analysis of nestedness from a functional perspective (i.e. a nestedness analysis which incorporates how similar species are in terms of their ecological traits). We use two null models to test the significance of any observed functional nestedness, and investigate the role of habitat island area in driving functional nestedness. We also determine whether functional nestedness is driven primarily by species composition or by differences in species' traits. We found that the majority (94%) of datasets were functionally nested by island area when a permutation null model was used, although only 11-22% of datasets were significantly functionally nested when a more conservative fixed-fixed null model was used. Species composition was always the most important driver of functional nestedness, but the effect of differences in species traits was occasionally quite large. Our results isolate the importance of island area in driving functional nestedness where it does occur, and the ordered loss of

functional traits due to habitat loss. This analysis demonstrates the potential insights that may derive from testing for ordered patterns of functional diversity.

9.3 Introduction

The fragmentation and destruction of natural habitat is the main driver of current biodiversity loss (Sala *et al.,* 2000; Whittaker & Fernández-Palacios, 2007). Understanding how species are affected by habitat loss is thus crucial to designing appropriate conservation and management responses (Fischer & Lindenmayer, 2007). The majority of habitat loss research, and ecological research more generally, has tended to focus solely on species identities (McGill *et al.,* 2006). However, it is increasingly acknowledged that functional diversity (i.e. the "extent of functional differences among the species in a community", Petchey & Gaston, 2002, p.402) is an important component of biodiversity and may provide valuable insight into ecosystem processes (e.g. Petchey & Gaston 2002, 2006; McGill *et al.*, 2006; Naeem *et al.,* 2012). It is also becoming apparent that, alongside species diversity, functional diversity is being eroded due to habitat loss (Flynn *et al.,* 2009; Naeem *et al.,* 2012; Newbold *et al.*, 2013). As a result, it has been argued that researchers should no longer focus solely on species identities and that there is a need for greater integration of functional diversity within traditional ecological subjects (e.g. McGill *et al.,* 2006; Devictor *et al.,* 2010; Safi, 2011; Monnet *et al.,* 2014). A focus on functional diversity allows workers to consider the differences between species in terms of their ecological traits, which in turn enable greater understanding of ecosystem functioning (Naeem *et al.,* 2012; Monnet *et al.,* 2014). For instance, maintaining functional diversity of birds in fragmented landscapes is important as birds provide numerous important functional roles, including pollination, the dispersal of seeds between fragments, and the predation of invertebrates and small mammals within fragments (Henle *et al.,* 2004; Tscharntke *et al.,* 2008; Dehling *et al.,* 2014).

Anthropogenic habitat destruction generally produces islands of natural habitat set in a matrix of disturbed landscape. Hence, ecologists have often used theories derived from the field of island biogeography to study such landscapes (Whittaker & Fernández-Palacios 2007). One widely studied insular phenomenon is nestedness, the analysis of which was formally introduced by Patterson & Atmar (1986). It can be defined as "a form of ordered composition of species assemblages involving a significant tendency for packing of the [presence–absence] matrix into a series of proper subsets" (Matthews *et al., in press*; and see Table 9.1). Examining the prevalence of nestedness in habitat islands is important from a conservation perspective as it provides information on the distribution of species in fragmented landscapes; information which can be used to establish conservation prescriptions, such as protected area placement (Whittaker & Fernández-Palacios, 2007, Triantis & Bhagwat, 2011).

Early studies found significant nestedness to be a near universal finding. For example, Watling & Donnelly (2006) reported that 94% of the 67 island datasets included in their synthetic analysis were significantly nested. However, recent studies have questioned the choice of nestedness indices and null models used in many of these earlier studies, claiming that the methods suffer from inflated type I errors (e.g. Ulrich & Gotelli, 2007). Analyses using more appropriate null models have found significant nestedness to be evident in only a minority of island datasets (e.g. Ulrich & Gotelli 2007, 2012; Matthews *et al., in press*). These methodological issues led to the development of more appropriate indices, such as a nestedness index based on overlap and decreasing fill (NODF; Almeida-Neto *et al.,* 2008; Table 9.1). NODF is based on the twin properties of standardized differences in presence–absence matrix row and column fills and paired overlap (i.e. the overlap of presences in two adjacent columns), and is generally considered the most statistically appropriate index available (Almeida-Neto *et al.,* 2008; Ulrich *et al.,* 2009).

Table 9.1 A glossary of terms used in this chapter.

Term (with abbreviation)	Description
A nestedness index based on overlap and decreasing fill (NODF)	NODF is a nestedness index based on the twin properties of standardized differences in presence–absence matrix row and column fills and paired overlap (i.e. the overlap of presences in two adjacent columns).
Fixed-fixed null model (FF)	A null model which creates a randomized presence–absence matrix that preserves row and column totals.
Functional nestedness	Used in this study to refer to the type of nestedness measured by traitNODF (below).
Nestedness (taxonomic)	A form of ordered composition of species assemblages involving a significant tendency for packing of the matrix into a series of proper subsets. In classic richness-ordered nestedness, the species composition of species-poor assemblages is a nested subset of the species composition of species-rich assemblages.
Permutation test null model (permRows)	A method for determining whether an observed index value is significantly different than expected by chance, whereby the rows of the species–site matrix are randomly re-ordered. The relevant index is calculated using this re-ordered matrix, and the process repeated *n* times to produce a distribution of null model values. permRows works best when then the original presence–absence matrix is organised according to an *a priori* hypothesis.
S.Fraction	The fraction of treeNODF that would be observed if all species were equally related.
topoNODF	The fraction of the treeNODF value due to the topology of the tree, i.e. the functional differences between species.
traitNODF	The variant of the treeNODF index which measures the nestedness of communities, taking into account how species are related in terms of their traits. The relationship between species in terms of their traits is determined by the construction of a dendrogram from the trait data.
treeNODF	An extension of the NODF index "in which the resemblance of descriptors, represented by a tree-like object, is taken into account. This can be carried out simply by replacing the incidence index (e.g. species richness of sites, frequency of occurrence of species) by the total branch length of descriptor variables" (Melo *et al.,* 2014, p. 564). treeNODF can be generalised for use with any tree-like object, such as a phylogenetic tree, or a functional dendrogram constructed using species traits. treeNODF can be partitioned into two values: S.Fraction and topoNODF (see above).
Z-transformed score (Z)	A method for standardising observed index values, using the formula: (Obs-μ)/sd. Where Obs is the observed nestedness value according to a given index, μ is the mean nestedness index value of the *n* null model permutations and sd is the standard deviation of the *n* values.

To date, studies of nestedness have focused exclusively on nestedness of species composition data. This traditional approach overlooks the fact that species which share the majority of their ecological characteristics will often be functionally redundant in a given ecosystem (Blackburn *et al.,* 2005; Melo *et al.,* 2014). Recently, Melo *et al.,* (2014) introduced an adaptation of the standard NODF index, termed treeNODF, in which the branch lengths of tree structures are incorporated, e.g. functional dendrograms and phylogenies. The variant of the treeNODF index which incorporates functional diversity was termed traitNODF, and it can be used to test for nestedness in ecological samples, whilst taking into consideration how similar species are in terms of their ecological traits ('herein functional nestedness'; see Fig. 9.1 for an example; and see Table 9.1 for a glossary of terms).

The treeNODF index is calculated in a similar manner to standard NODF, but instead of using the sum of species occurrences to estimate the degree of nestedness, the branch lengths of a tree-like object are used (i.e. species richness/species incidence is replaced with the branch lengths of a tree-like object). In traitNODF, the tree-like object is a functional dendrogram constructed using species trait data, whereby the branch lengths of this dendrogram represent the functional diversity (FD) of the system in question (e.g. Fig. 9.1). traitNODF works by first ordering the sites along a particular gradient (e.g. in this example, island area). All pairwise combinations of sites along this area gradient are then calculated. A pair of sites is then taken (A & B), whereby it is initially assumed that as A is larger than B, A is a relatively high FD community and B is a relatively low FD community. The FD of the two communities is then compared. If the FD of A is found to be lower than the FD of B, traitNODF takes a value of zero. If the FD of A is greater than the FD of B, the traitNODF value of these two sites is the proportion of the branches in the dendrogram of B that are also present in the dendrogram of A (Melo *et al.,* 2014). As Melo *et al.* state, in the case of using

treeNODF with a phylogenetic tree, if the branch lengths of the phylogenetic tree are considered proportional to the number of features, then the treeNODF value of these two sites is the proportion of the features present in the features-poor site that are also present in the features-rich site. This pairwise comparison of sites is repeated across all sites in the dataset and the resultant traitNODF values are then averaged to give a traitNODF value for all sites.

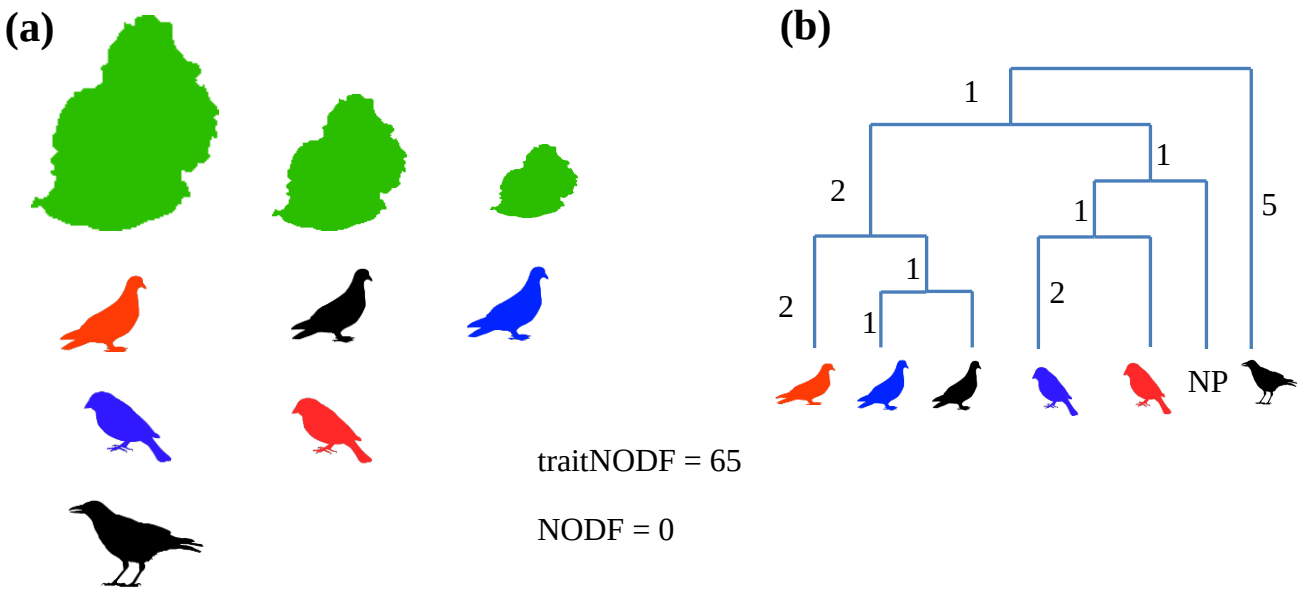


Figure 9.1 A hypothetical example illustrating the concept of functional nestedness, and in particular the situation where a set of islands can be functionally nested but not taxonomically nested. In (a), a number of different bird species are distributed across a set of islands of varying size (green shapes), such that the largest island has the most species. There are six different species of birds in the archipelago, each represented by a different colored symbol. The different colored versions of the same bird symbol indicate species which have only recently diverged, and are thus functionally similar. The exact relationship between these species in terms of their traits is given in the tree-object in (b). The numbers in (b) represent the branch lengths. It can be seen that the functionally similar species (i.e. the different colored versions of the same symbols) share most of their branch lengths, but not all. Using A and B it is then possible to calculate both traitNODF and the standard NODF index values, which are 65 and 0 respectively. Standard NODF is zero because this index only focuses on species incidences (i.e. not the functional similarity of species), and none of the species occur on more than one island (A). NP = a species not present in this archipelago. The bird outlines were provided by Susan Spear, courtesy of the Cornell Lab of Ornithology.

Melo *et al.* (2014) illustrated that using traitNODF can be enlightening as standard NODF and traitNODF may show different patterns; that is, a set of isolates may be

significantly nested in terms of functional traits but not in terms of species composition (e.g. see Fig. 9.1) and *vice versa*. A useful feature of treeNODF (and by extension, traitNODF) is that it can be partitioned into two components: 1) S.Fraction – the composition component of the index, i.e. the fraction of treeNODF that would be observed if all species were equally related, and 2) topoNODF – the fraction of the treeNODF value due to the topology of the tree (e.g. the dendrogram). Examining each component separately allows for additional inferences regarding the nestedness of ecological communities (Melo *et al.*, 2014). Furthermore, functional nestedness potentially offers an explanation for taxonomic nestedness in certain instances as it provides a link between species' distributions and the environment.

When testing for nestedness, a sensible approach is to determine an *a priori* hypothesis regarding the particular mechanism(s) likely to be driving any observed nestedness, rather than simply organizing the presence–absence matrix to achieve the maximum possible nestedness value (McAbendroth *et al.,* 2005; Almeida-Neto *et al.,* 2008; Melo *et al.,* 2014). For example, in both true island and habitat island systems island area (i.e. differential area requirements between species) and isolation (i.e. differential dispersal abilities between species) have often been found to be important drivers of both nestedness and functional diversity (e.g. Fig. 9.1; Patterson & Atmar 1986*;* McAbendroth *et al.,* 2005; Ulrich *et al.,* 2009; Ding *et al.,* 2013; Whittaker *et al.,* 2014; Matthews *et al., in press*). Thus, in such systems a more fruitful approach to maximally packing the presence–absence matrix is to order sites according to island area (isolation, disturbance intensity, etc., could just as easily be used) and then compute the nestedness index, keeping these site orders fixed. Using this approach Melo *et al.* (2014) found that island area was a good determinant of functional nestedness for a Caribbean bat community.

To the best of our knowledge, traitNODF has yet to be used to determine the prevalence of functional nestedness in ecological communities, with the exception of Melo *et al.*'s (2014) single study of Caribbean bat communities. Here, we analyse functional nestedness patterns in 18 bird–forest habitat islands datasets spanning a broad array of ecological contexts. To be clear, in this study we are interested in functional and compositional nestedness between sites. It is possible to calculate overall functional nestedness (i.e. nestedness between sites and species) but it requires site-specific environmental data which were generally lacking from the source papers. We focus on birds because (a) it possible to obtain functional trait data for most species of interest, and (b) birds are known to perform essential ecological functions in forests, including seed dispersal and predation on insect herbivores (Bregman *et al.,* 2014). We use a comprehensive and well-specified database of bird traits (see Tobias *et al.,* 2014), containing eight continuous traits (e.g. bill depth, bill width, tarsus length) to build functional dendrograms for over 1000 bird species in our datasets.

Our study has two primary aims. First, we assess whether habitat islands are more or less functionally nested by island area than expected by chance (based on two null models). Our analyses are based on the idea of ordering presence-absence matrices according to an *a priori* hypothesis (i.e. by island area), and we do not test for maximally packed nestedness values. Based on previous studies which have found: (a) evidence of trait-mediated loss of bird species due to habitat loss (Newbold *et al.,* 2013; Bregman *et al.,* 2014), and (b) significant species–area relationships in fragmented landscapes (Lees & Peres, 2008; Matthews *et al.,* 2014a), we hypothesize that a large number of our datasets will be significantly functionally nested by island area. Second, by partitioning treeNODF values into their S.Fraction and topoNODF components we aim to determine whether observed functional nestedness is driven primarily by compositional nestedness or the tree topology.

We discuss the implications of our results in the context of ecosystem functioning and conservation biogeography.

9.4 Methods

9.4.1 Dataset collection

Habitat island datasets were sourced using several abstracting databases between May 2010 and August 2014, and a variety of keywords, including 'habitat islands', 'species richness', and 'fragments'. Certain datasets were obtained from the authors of the source papers, whilst others were supplemented with additional data from the source paper authors. Following Matthews *et al.* (2014), suitable datasets were screened based on the following criteria:

1. habitat islands were defined as discrete patches of forest surrounded by contrasting matrix habitat;
2. there were at least six habitat islands;
3. a list of bird species in each habitat island within the dataset was given;
4. it was possible to obtain trait data for nearly all (ideally 100%; for details see below) of species within a dataset; and

9.4.2 Trait data

In total, we collected data on eight continuous traits (the method for trait data collection is outlined in section 3.7 of this thesis. As individual traits are highly variable (e.g. Trisos *et al.*, 2014), and many of the traits exhibit colinearity, we condensed these trait data into a number of indices. First, following Trisos *et al.*, (2014) we conducted two Principal Components Analyses (PCA): one using the bill shape measurements ('A'; culmen length, width, length from nares, depth), and one using the locomotive measurements ('B'; tarsus length, wing chord, and tail length). The second axes of A (short beak:wide/deep beak) and B (short tarsus:long tail/wing) were then taken as a bill shape index, and a locomotive index, respectively. A further PCA analysis ('C') was then undertaken using the first axes of A and

B as the input; the first axis of C was then taken as a body size index (Trisos *et al.,* 2014). Finally, we created a dispersal index standardizing for bird size, using the hand–wing index (Claramunt *et al.,* 2012). This index was calculated using the formula: (*K*/*W*) * 100, where *K* is the Kipp's Distance, and *W* is the wing chord length (Claramunt *et al.,* 2012). In sum, the traits used in the functional nestedness analyses were the body size index, the locomotive index, the dispersal index, and the beak shape index. All traits were standardized to have a mean of zero and a standard deviation of one (Mouchet *et al.,* 2008) using the scale function in R. These standardized traits were then used to build functional dendrograms (Petchey & Gaston, 2002) for each dataset. First, the species-by-trait matrix for a given dataset was converted into a distance matrix using Euclidean distance (Petchey *et al.,* 2004). Second, the distance matrix was subject to a cluster analysis (UPGMA method) to create a dendrogram (Petchey & Gaston, 2002; Petchey *et al.,* 2004). The distance matrix was constructed using the 'vegan' package (Oksanen *et al.,* 2013), the cluster analysis was performed in the base stats R package, and the PCA analyses were performed using the 'ade4' R package (Dray & Dufour, 2007).

9.4.3 Measuring functional nestedness

The treeNODF index works by assessing the branch lengths in a tree-like object. Thus, each dendrogram (i.e. the dendrogram for each dataset) was converted into a tree object using the 'as.phylo' function in the 'ape' R package (Paradis *et al.,* 2004; and see our Fig. 9.2). The site–species matrices were ordered according to island area, and the row orders fixed. The traitNODF index (for rows, i.e. sites) was then calculated for each area-ordered dataset using the dataset's sites–species matrix and corresponding dendrogram, and the 'CommEcol' R package (Melo, 2012).

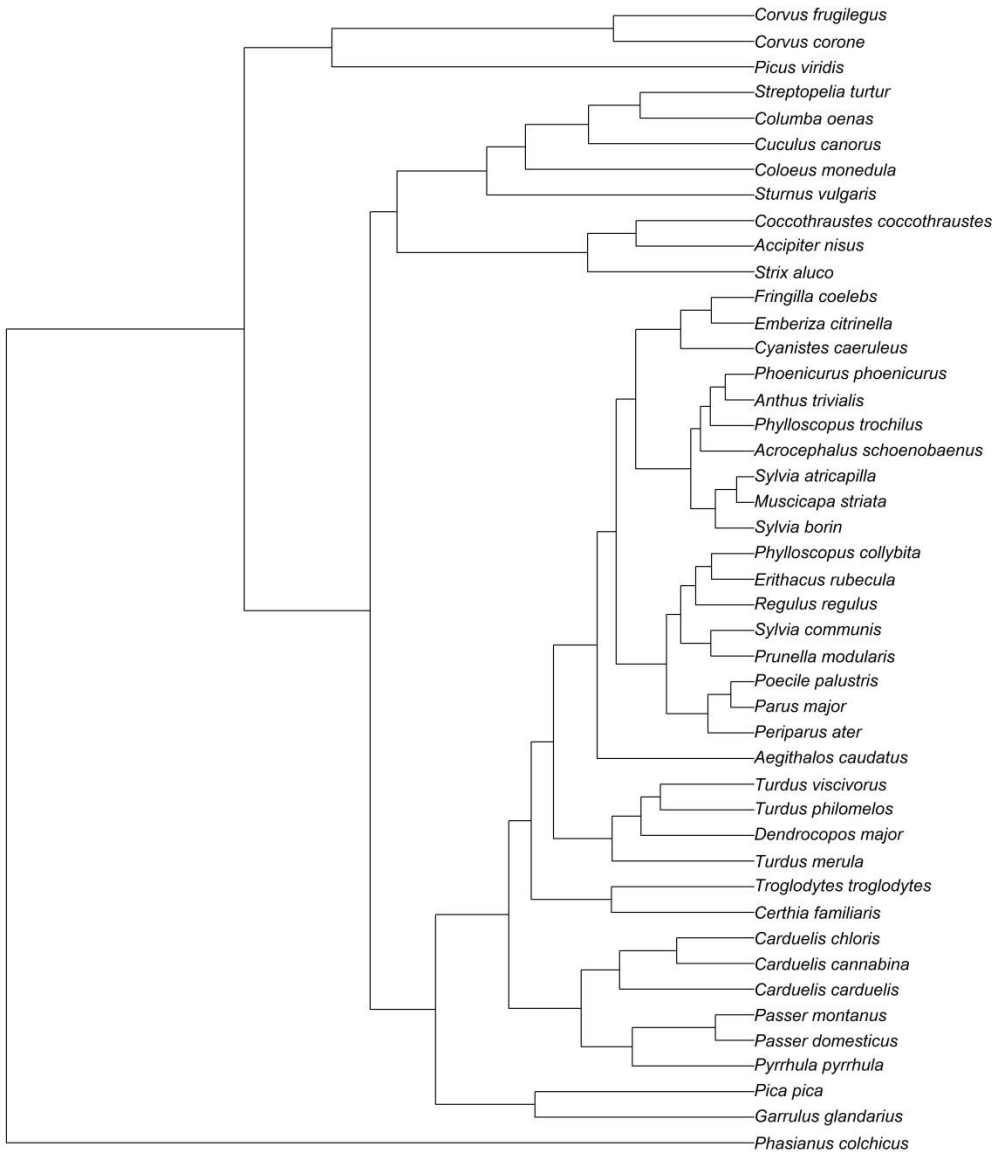


Figure 9.2 An exemplar dendrogram converted into a tree object, to be used with traitNODF. The data are from McCollin (1993; see Table A7.1 in Appendix A7). The dendrogram was constructed by first transforming the species–trait data into a distance matrix (Euclidean distance), and then using the UPGMA clustering algorithm. The resulting dendrogram was then converted into a tree object (see 'Methods').

In addition to storing the traitNODF value for each dataset, we also stored the partition values: S.Fraction and topoNODF (see Table 9.1). As S.Fraction is the compositional component of traitNODF it is generally similar to the standard NODF index.

However, as standard NODF and S.Fraction are not necessarily identical (A.S. Melo *Personal communication*), we also computed the standard NODF index (again taking the value for rows) for each area-ordered dataset using the vegan R package.

To determine whether the observed traitNODF, S.Fraction, topoNODF and standard NODF values differed from those expected by chance we first used a permutation test null model approach ('permRows'; Melo *et al.,* 2014). To implement permRows, we used the 'treeNodftest' function in the CommEcol package. This function randomly permuted the rows of the species–sites matrix (i.e. permuted the sites) and calculated the relevant index (e.g. traitNODF, standard NODF) using this permuted matrix. The matrix was then re-permuted and the processes repeated 1000 times, for each dataset. As whole rows are permuted, this null model preserves the number of sites used by each species (i.e. fixed column totals), and the species and functional composition of sites is maintained. Using the permRows test is useful when sites are ordered according to an *a priori* hypothesis and you are interested in determining the mechanisms driving any observed nestedness. We took the *P* value generated by the treeNodftest function and employed an α level of 0.05. It is important to note that, a "non-significant" result using permRows does not mean that the matrix is not significantly nested, but that the site ordering hypothesis lacks explanatory power for nestedness in the matrix. We also calculated the Z-transformed score for each dataset, where $Z= \frac{Obs-\mu}{sd}$, and where Obs is the observed nestedness value according to a given index, μ is the mean nestedness index value of the null model values, and sd is the standard deviation of the 1000 null model iterations.

As the choice of null model has been found to influence results in nestedness studies (Ulrich & Gotelli 2007, 2012), we also ran the more conservative fixed-fixed null model (FF). The FF model creates a randomized presence–absence matrix of the same dimensions

as the observed matrix and keeps marginal row and column totals fixed. We ran the FF model to determine the significance of traitNODF values for the area-ordered matrices (i.e. row orders were fixed). We had originally also planned to test the effect of isolation on functional nestedness by running the two null model analysing using isolation-ordered matrices, for a subset of datasets for which isolation data were available. However, as only a few studies were found to have measured isolation and these studies all used different isolation metrics, the results were difficult to interpret. Thus, we decided to focus solely on the role island area.

9.4.4 Sensitivity analyses

The choice of traits to use in functional diversity studies is important and it can be useful to determine how robust the results of analyses are to the removal of individual traits (Blackburn *et al.*, 2005; Petchey and Gaston 2006; Whittaker *et al.,* 2014). To this end we undertook a sensitivity analysis whereby we removed a trait, re-ran the permRows analyses and recorded the results. The removed trait was then re-instated and a different trait removed, and so on until the analyses had been run with each of the traits lacking. Analysis of these results allowed us to determine whether our findings were robust to the removal of a particular trait. In this context, a trait refers to one of the indices (i.e. a combination of multiple traits). Removing an actual individual trait (e.g. tail length) would have the same effect as it would mean an index could not be computed. All analyses were conducted in R (version 3.0.2; R Development Core Team 2013).

9.5 Results

9.5.1 The nature of the datasets

We found 18 datasets that matched criteria 1–4 (see Table A7.1 in Appendix 7). In order to calculate traitNODF, we removed islands with zero species from the relevant datasets. It was also necessary to remove species from datasets for which we were unable to acquire the trait

measurements. This occurred, for example, because the species were endangered and thus specimens were hard to locate. In practice, for 7 of the 18 datasets we had complete coverage of species (i.e. we obtained trait data for 100% of species). For 4 datasets we were missing data on 1 species, 3 datasets had 2 species without trait data, and for 4 datasets we were missing data for between 3 and 8 species (the relevant datasets are highlighted in Table S1). As species with missing data represented a very small proportion of the total number of species within a dataset, we consider it unlikely that these 'missing species' affected our results.

9.5.2 Functional nestedness by island area

When the presence–absence matrices were ordered according to island area the traitNODF values were relatively high, i.e. closer to 100 (maximal nestedness) than 0 (mean traitNODF for area ordered matrices = 71.9). When a permutation test (1000 permutations) was used to test for the significance of the traitNODF values, the observed traitNODF value of the area-ordered matrix was significantly larger than expected by chance in 17 of the 18 datasets, indicating that for these datasets the area-ordered matrix was more functionally nested than the majority of other matrix order permutations (Table 2). The Z-transformed scores for the area-ordered datasets were all positive and were relatively high (mean Z-transformed score = 4.0).

Partitioning the area traitNODF value into its two components (S.Fraction and topoNODF) revealed that, in general, the traitNODF value of a dataset was primarily comprised of the S.Fraction component (i.e. the species composition component), rather than the topoNODF component (i.e. the tree topology component; Table 9.1). topoNODF accounted for between 4.3% to 35.4% of total traitNODF (mean = 13.0%). In all but one dataset, the permutation test results were consistent between traitNODF, S.Fraction and topoNODF; that is, if the traitNODF value was significant for a dataset, so were the

S.Fraction and topoNODF values. For a single dataset (8) the traitNODF and S.Fraction values were significant, whilst the topoNODF value was not (Table 2).

As expected, when the FF model was used, significant functional nestedness was much less prevalent (Table A7.2 in Appendix 7). Two datasets had a significant traitNODF value and a positive Z-score, whilst four datasets had a significant topoNODF value and a positive Z-score. No dataset had a significant S.Fraction value. A number of the datasets had negative Z-scores, although none of these were significant (Table A7.2).

As the standard NODF index and S.Fraction may not be equivalent we calculated both. The standard NODF values were very similar to the S.Fraction values for each dataset (Pearson's correlation coefficient = 0.93, $P = < 0.001$), and thus for these datasets S.Fraction and standard NODF are roughly interchangeable. Furthermore, the permutation test results for traitNODF generally followed the standard NODF permutation test results. Indeed, for 16 datasets, both traitNODF and standard NODF produced the same results. For two datasets, standard NODF was not significant, whilst traitNODF was (Table 2).

Table 9.2 traitNODF and standard NODF values (by rows) for the area ordered matrices of 18 bird-habitat island datasets. Prior to calculation of the nestedness indices, for each dataset the presence–absence matrix was ordered according to habitat island area, and the row (site) orders fixed. traitNODF (Trait) values were partitioned into their two components: S.Fraction (S.Frac.) and topoNODF(Topo). The standard NODF (Standard) index (by rows) was also calculated for each area-ordered matrix. For each of the four measures (traitNODF, S.Fraction, topoNODF, and standard NODF) significance was determine using a permutation test (1000 permutations), and a *P* value generated from the resultant index values. From these permutations, a Z-transformed score was also generated for traitNODF (Z). The proportion of traitNODF represented by topoNODF (Topo Prop.) is also presented. The dataset numbers correspond to those in Table A7.1.

No.	Trait (NODF)	Trait (*P*)	Trait (Z)	Topo (NODF)	Topo (*P*)	S.Frac. (NODF)	S.Frac. (*P*)	Standard (NODF)	Standard (*P*)	Topo Prop.
1	95.38	<0.01	4.25	4.08	<0.01	91.29	<0.01	90.03	<0.01	4.28
2	61.89	<0.01	4.65	8.03	<0.01	53.87	<0.01	53.89	<0.01	12.97
3	77.33	0.03	2.14	9.65	0.04	67.68	0.02	62.86	0.06	12.48
4	74.53	<0.01	7.48	16.25	<0.01	58.29	<0.01	58.28	<0.01	21.80
5	75.12	<0.01	3.40	8.85	<0.01	66.27	<0.01	66.58	<0.01	11.78
6	69.35	<0.01	3.44	5.54	<0.01	63.82	<0.01	60.98	<0.01	7.99
7	68.04	<0.01	3.44	4.20	0.02	63.83	<0.01	62.17	<0.01	6.17
8	88.04	<0.01	3.69	5.86	0.05	82.18	<0.01	81.90	<0.01	6.66
9	52.63	0.24	0.75	6.32	0.43	46.32	0.21	55.32	0.11	12.01
10	69.87	0.01	2.12	8.35	0.03	61.52	0.01	65.49	0.02	11.95
11	70.89	<0.01	5.43	12.29	<0.01	58.59	<0.01	61.00	<0.01	17.34
12	71.99	<0.01	6.36	11.27	<0.01	60.71	<0.01	61.68	<0.01	15.65
13	64.69	<0.01	4.92	9.40	<0.01	55.29	<0.01	54.26	<0.01	14.53
14	70.34	<0.01	5.33	7.60	<0.01	62.74	<0.01	64.25	<0.01	10.80
15	63.90	<0.01	2.96	5.70	<0.01	58.20	<0.01	55.35	<0.01	8.92
16	78.42	<0.01	5.72	6.65	<0.01	71.78	<0.01	72.75	<0.01	8.48
17	69.43	<0.01	3.73	24.57	<0.01	44.86	<0.01	42.72	<0.01	35.39
18	73.21	<0.01	2.85	10.8	<0.01	62.41	<0.01	50.79	0.05	14.75

9.5.3 Sensitivity analyses

The island area-ordered matrix results were robust to the removal of any single trait/index. In all datasets the significance of the observed traitNODF value did not change in any of the iterations (i.e. any of the analyses conducted after removing a trait; Table 3). Similarly, the actual traitNODF values varied only slightly between iterations (Table 3).

Table 9.3 TraitNODF area results from the sensitivity analysis. For the sensitivity analysis we first calculated traitNODF for each dataset using all four trait indices. We then removed a trait index, and re-ran the analyses. This trait index was then re-instated and the next trait index removed, and so on.

No.	All		Trait Removed							
			Trait 1		Trait 2		Trait 3		Trait 4	
	tN	*P*	tN	*P*	tN	*P*	tN	*P*	tN	*P*
1	95.4	<0.01	96.2	<0.01	95.9	<0.01	93.7	<0.01	94.6	<0.01
2	61.9	<0.01	63.7	<0.01	63.0	<0.01	61.3	<0.01	62.9	<0.01
3	77.3	0.03	78.9	0.02	77.8	0.03	78.9	0.02	78.3	0.02
4	74.5	<0.01	76.5	<0.01	75.0	<0.01	77.8	<0.01	74.8	<0.01
5	75.1	<0.01	71.1	<0.01	79.2	<0.01	71.7	<0.01	72.8	<0.01
6	69.4	<0.01	70.7	<0.01	68.9	<0.01	70.5	<0.01	69.1	<0.01
7	68.0	<0.01	68.7	<0.01	68.8	<0.01	71.2	<0.01	71.2	<0.01
8	88.0	<0.01	87.7	<0.01	86.6	<0.01	90.1	<0.01	88.2	<0.01
9	52.6	0.24	61.4	0.11	62.0	0.11	52.9	0.21	53.8	0.24
10	69.9	0.01	71.0	0.02	77.3	0.01	67.9	0.04	71.2	0.01
11	70.9	<0.01	68.0	<0.01	70.4	<0.01	73.2	<0.01	69.0	<0.01
12	72.0	<0.01	73.6	<0.01	71.6	<0.01	72.6	<0.01	74.3	<0.01
13	64.7	<0.01	67.4	<0.01	65.1	<0.01	66.9	<0.01	63.7	<0.01
14	70.3	<0.01	71.4	<0.01	74.2	<0.01	70.7	<0.01	71.6	<0.01
15	63.9	<0.01	65.2	<0.01	68.1	<0.01	64.3	<0.01	60.0	0.01
16	78.4	<0.01	79.4	<0.01	77.9	<0.01	79.5	<0.01	79.3	<0.01
17	69.4	<0.01	71.6	<0.01	71.6	<0.01	68.4	<0.01	72.7	<0.01
18	73.2	<0.01	67.3	0.01	78.7	<0.01	72.9	<0.01	70.1	<0.01

9.6 Discussion

We have undertaken a synthetic analysis of functional nestedness patterns of birds in 18 habitat island datasets. Our results were sensitive to the null model approach used. When a permutation test was used area-driven functional nestedness was found to be relatively common. However, when the more conservative FF model was used, most datasets were not significantly functionally nested. In general, the observed functional nestedness of datasets was more a result of the species composition of islands than the functional differences between species. That being said, topoNODF represented a sizeable proportion of traitNODF for a number of datasets, indicating that the inclusion of functional traits into nestedness analyses can be revealing in certain instances.

9.6.1 The prevalence of functional nestedness in habitat islands

With a single exception (Melo *et al.,* 2014), the study of nestedness in islands has focused exclusively on the nestedness of species composition. The development of treeNODF allows nestedness theory to be applied to other important components of biodiversity, including functional diversity. This is important because examining the different functional roles species undertake in ecological communities is essential if we are to understand how communities are structured (Flynn *et al.*, 2009; Safi, 2011). To this end, the search for broad-scale patterns in functional diversity, such as functional nestedness, represents a critical frontier in this field (Petchey & Gaston, 2002; Flynn *et al.*, 2009).

It is not possible to directly compare the results of our analyses with previous synthetic analyses of nestedness patterns in habitat islands (e.g. Watling & Donnelly, 2006; Matthews *et al., in press*), as we have adopted a different methodological approach. Whereas previous studies have simply organized the presence–absence matrix to achieve the maximum nestedness value, we have ordered the matrix by island area, and our results (e.g. treeNODF and standard NODF values) relate to these matrices.

Using a permutation algorithm we found that when the presence–absence matrices were ordered according to island area, the majority of datasets were functionally nested by area. In almost all cases where the traitNODF value was significantly different from that expected by chance according to permRows, the S.Fraction and topoNODF components were also significant. Thus, based on these results, one could infer that in these datasets small islands do not simply contain species assemblages that are a subset of the larger islands, but the small islands also contain a number of functionally similar species (i.e. species which share a large proportion of their dendrogram branch lengths; see Fig. 9.1) to the larger islands. However, as to be expected, when FF was used, a much smaller proportion of datasets were significantly functionally nested. Whilst the two null models have generated

different results, it is difficult to directly compare the results as permRows and FF represent different approaches to null modelling; that is, using each is to ask a slightly different question regarding significance. Our permRows results indicate that the traitNODF values generated from the area-ordered matrices are larger than most other row-ordered permutations of the matrix. This is why permRows has been argued to represent a useful method when ordering matrices according to an *a priori* hypothesis. In contrast, our FF results indicate that the traitNODF values were generally not significantly different from the values expected after the presences in the area-ordered matrices had been randomized; with the constraint that marginal totals are kept constant. Thus, FF differs in several ways from permRows, and is known to be a very conservative null model (Ulrich *et al.*, 2009; Strona & Fattorini, 2014). Taking the two sets of results together, a tentative conclusion would be that our area-ordered matrices are generally not more functionally nested than expected according to a randomly ordered-matrix, but that island area provides a reasonable explanation for what functional nestedness is present.

Partitioning the observed traitNODF values into their constituent components (S.Fraction and topoNODF; see Table 9.1) indicated that the S.Fraction component (i.e. the compositional component of traitNODF) always comprised a greater proportion of traitNODF than the topoNODF component. In Melo *et al.*'s (2014) analysis of Caribbean bat communities, topoNODF also made up a smaller proportion of traitNODF (~29%) than S.Fraction. Nonetheless, in our study topoNODF was occasionally quite large (e.g. 35% for dataset 17; Table 2), and with one exception was always significant when traitNODF was significant, when permRows was used. As the number of datasets involved in the present study is relatively small, further tests for functional nestedness in habitat islands will be necessary to determine the generality of our findings.

That topoNODF often comprised a low fraction of traitNODF could potentially be a result of us not selecting the most appropriate traits. The choice of traits to use in functional diversity studies can be critical (Petchey & Gaston, 2006; Mouchet *et al.,* 2008), and given the complex nature of ecological communities it is relatively easy to (a) overlook important traits, and/or (b) incorporate uninformative traits (Cadotte *et al.,* 2013). Our sensitivity analyses and use of composite trait indices makes the second of these possibilities unlikely, but the first possibility is harder to rule out. However, we have obtained continuous trait data relating to eight different measurements that are known to characterize the functional roles of bird species (Henle *et al.,* 2004; Sutherland *et al.,* 2004; Tobias *et al.,* 2014; Trisos *et al.,* 2014), and thus we are confident that the key traits have been incorporated and quantified.

9.6.2 Driver of functional nestedness

Rather than simply ordering the presence–absence matrices to obtain the maximum possible traitNODF value, we have based our analyses on an *a priori* hypothesis: that island area will underpin observed functional nestedness patterns. Area has been postulated to drive nestedness due to area-dependent extinctions, and species are known to have varying minimum area requirements, such that large patches contain species with small and large minimum area requirements, whilst smaller patches only contain species with small area requirements (Patterson & Atmar, 1986; Matthews *et al., in press*). For example, the area of ponds was found to drive nestedness patterns of macroinvertebrates in the UK (McAbendroth *et al.,* 2005). From a functional perspective, forest specialist species possess traits (e.g. large body size, low dispersal ability) which demand larger area requirements relative to taxonomically similar generalist species (Henle *et al.,* 2004; Matthews *et al.,* 2014a). The island area results provide more evidence for the importance of habitat island area in maintaining species diversity in fragmented landscapes (discussed below; Whittaker & Fernández-Palacios, 2007; Matthews *et al.,* 2014a).

9.6.3 Implications for conservation and ecosystem functioning

Until recently, the majority of conservation theory has been based on analyses of taxonomic diversity. However, a number of studies have shown that taxonomic diversity does not necessarily have high congruence with other facets of diversity, including functional diversity (e.g. Monnet *et al.,* 2014). One major issue with purely compositional approaches is that they do not allow inferences to be made regarding the effects of species loss on ecosystem functioning (Ding *et al.,* 2013) – an important issue given that the loss of species with particular traits can impact the delivery of particular ecosystem functions, depending on the level of functional redundancy (Newbold *et al.,* 2013). Studies that have examined the effects of anthropogenic environmental change on functional diversity have found that functional diversity often declines in human dominated landscapes at a greater rate than species richness (Flynn *et al.,* 2009; Ding *et al.,* 2013, Boyer & Jetz, 2014). For example, up to 80% of the original functional diversity of bird communities has been lost from certain oceanic islands due to pre-historical and historical extinctions (Boyer & Jetz, 2014). Several authors have argued, on the basis of such findings, that functional diversity should be more explicitly considered in biodiversity conservation, for instance in the location of protected areas (e.g. Naeem *et al.,* 2012; Devictor *et al.,* 2014). In this regard, functional nestedness represents a potentially useful index for conservation planners looking to account for broad scale patterns of functional diversity.

There are linkages between the concept of functional nestedness and Diamond's (1975b) assembly rules (reviewed in Whittaker & Fernández-Palacios, 2007). Diamond (1975b) argued that the distribution of bird species across islands in the Bismark Archipelago followed a set of rules mediated by interspecific competition; new species could only enter an island/community in accord with these assembly rules. In relation to functional nestedness, if we observe functional nestedness by island area, it suggests that there is a particular sequence

in which different types of species are added to a system as it increases in size. However, whilst Diamond's assembly rules were largely based on interspecific competition, functional nestedness may occur for a number of reasons, such as area-dependent loss of trait space (discussed above). To take a simple hypothetical example, if a set of fragments are functionally nested in terms of feeding guild, such that the loss of guilds occurs in an ordered manner according to island area, one might envisage insectivores being the most area-sensitive species (e.g. Şekercioğlu *et al.,* 2002, found insectivore species to be particularly sensitive to forest loss in Costa Rica), with frugivores being less area-sensitive (Bregman *et al.,* 2014). In this example insectivores would be absent from the smallest islands and only occur in the largest islands. Thus, the valuable functional role of pest control and insect herbivory would be lost from smaller fragments. This ordered loss of guilds by area would also provide an explanation for the observation that this system was taxonomically nested. However, whilst our results do indicate at least a small degree of functional nestedness by area, as topoNODF often represented a low fraction of traitNODF, there must be an additional explanation for the observed taxonomic nestedness. Analysis of phylogenetic relationships between species in the datasets may prove enlightening in this regard (Melo *et al.,* 2014).

Researchers have used standard compositional nestedness indices for over twenty years to argue for specific conservation prescriptions, and in efforts to address the SLOSS debate (i.e. is a single large reserve more or less effective than several small reserves; see Whittaker & Fernández-Palacios, 2007, Matthews *et al., in press*). However, finding that an archipelago is or is not nested in species composition does not necessarily predict functional nestedness (Melo *et al.,* 2014). traitNODF does provide this information, and significant functional nestedness, as we have found in some of our datasets, indicates that the loss of traits from islands of decreasing area is proceeding in an ordered manner. That is, species

with unique traits are being lost in a systematic fashion – not perfectly, but to a significant degree (see also Flynn *et al.,* 2009). This is important because the order of species loss from a community can have substantial impacts on ecosystem functioning in that community (Petchey & Gaston, 2002). Nonetheless, as has been argued in the context of taxonomic nestedness, it is problematic to generate conservation guidance when a system is significantly functionally nested, but not perfectly functionally nested (Fischer and Lindenmayer 2005). Perfect functional nestedness in an area-ordered matrix would imply that by conserving the largest island in an archipelago, the total spectrum of species' traits would be conserved. As we never found perfect functional nestedness in our datasets, we do not recommend such an approach (i.e. conserving the largest island only) to be used for habitat island systems. However, our results do indicate that there is at least a small degree of ordered loss of bird traits in some of our datasets: smaller islands contain a number of functionally similar species to the species on larger islands. At the very least, this needs to be considered in future studies that attempt to estimate functional diversity loss resulting from habitat fragmentation.

9.6.4 Concluding remarks

Nestedness in island systems has been studied intensively over the last two decades, but these studies have been focused on species richness and composition and have ignored the importance of species traits in driving ordered patterns of biodiversity. Global environmental change continues to have a detrimental effect on all facets of diversity, including functional diversity (Flynn *et al.,* 2009; Devictor *et al.,* 2010). Further work is clearly needed to test for and evaluate the causes of ordered patterns of functional diversity (Naeem *et al.,* 2012). Functional nestedness analysis presents one such way to meet this challenge. We found that "significant" functional nestedness was relatively common when presence–absence matrices were ordered according to island area and a permutation null model approach was used, but was much less prevalent when the FF null model was used. Whilst acknowledging that our

results depend in part on the null model used, our findings are consistent with (but don't test explicitly) previous work that has found species traits to provide valuable information in modelling species at risk of extinction due to habitat loss (e.g. Şekercioğlu *et al.*, 2002, Bregman *et al.*, 2014). As such, we demonstrate that traitNODF can reveal interesting functional patterns that are masked when one focuses purely on species composition. Future studies should expand our analyses to test for functional nestedness using different taxa, and different types of study system (e.g. oceanic islands and food web networks).

CHAPTER 10

The gambin model provides a superior fit to species abundance distributions with a single free parameter: evidence, implementation and interpretation

Paper published in Ecography: Matthews, T.J., Borregaard, M.K., Ugland, K., Borges, P.A.V., Rigal, F., Cardoso, P. & Whittaker, R.J. (2014) The gambin model provides a superior fit to species abundance distributions with a single free parameter: evidence, implementation and interpretation. *Ecography*. **37**, 1002–1011.

10.1 Preface

This chapter is the seventh empirical chapter, and presents a new method and R package for fitting the gambin species abundance distribution model. The fit of the model is evaluated using a number of different datasets, and the utility of the model as an ecological diversity metric is examined. It was found that the model provided a very good fit to empirical data, and the model's single parameter provides a useful index of SAD form, which can be used in applied ecological contexts.

Authors: Thomas J. Matthews, Michael K. Borregaard, Karl I. Ugland, Paulo A.V. Borges, François Rigal, Pedro Cardoso, Robert J. Whittaker

The co-author contributions were as follows. T.J.M., K.I.U. and R.J.W. conceived the ideas; T.J.M. ran the analyses; and T.J.M. and R.J.W. wrote the paper with a significant contribution from the all authors.

10.2 Abstract

The species abundance distribution (SAD) has been a central focus of community ecology for over fifty years, and is currently the subject of widespread renewed interest. The gambin model has recently been proposed as a model that provides a superior fit to commonly preferred SAD models. It has also been argued that the model's single parameter (α) presents a potentially informative ecological diversity metric, because it summarises the shape of the SAD in a single number. Despite this potential, few empirical tests of the model have been undertaken, perhaps because the necessary methods and software for fitting the model have not existed. Here, we derive a maximum likelihood method to fit the model, and use it to undertake a comprehensive comparative analysis of the fit of the gambin model. The functions and computational code to fit the model are incorporated in a newly developed free-to-download R package (*gambin*). We test the gambin model using a variety of datasets and compare the fit of the gambin model to fits obtained using the Poisson lognormal, logseries and zero-sum multinomial distributions. We found that gambin almost universally provided a better fit to the data and that the fit was consistent for a variety of sample grain sizes. We demonstrate how α can be used to differentiate intelligibly between community structures of Azorean arthropods sampled in different land use types. We conclude that gambin presents a flexible model capable of fitting a wide variety of observed SAD data, while providing a useful index of SAD form in its single fitted parameter. As such, gambin has wide potential applicability in the study of SADs, and ecology more generally.

10.3 Introduction

A species abundance distribution (herein 'SAD') describes the abundances of all species sampled within a given community (Ulrich *et al.*, 2010). Because SADs characterise the structure of ecological communities, they constitute the foundation of much (macro-)

ecological and biogeographical theory (Preston, 1948). Thus, considerable effort has been made to characterise empirical SADs in a statistically tractable framework. To date around 30 different SAD models have been published (McGill *et al.,* 2007). The two most commonly used are the logseries (Fisher *et al.,* 1943) and the lognormal (Preston, 1948). The Poisson lognormal ('PLN'; Bulmer, 1974) is usually preferred to the continuous lognormal as the PLN incorporates a sampling theory and the continuous lognormal allows fractional abundances. The neutral model of Hubbell (2001) also predicts a SAD, termed the zero-sum multinomial distribution ('ZSM').

The logseries has often been found to provide a poor fit to empirical SADs (Ulrich *et al.,* 2010; but see White *et al.,* 2012), and the use of the PLN has been criticised, both in general terms (Williamson & Gaston, 2005), and particularly in comparative analysis, as the parameters are not as intuitively interpretable as those of the logseries (but see Engen, 2001, Sæther *et al.,* 2013). Thus, there is a need for a model that provides a good fit to a variety of empirical data, is analytically tractable and possesses an easily interpreted parameter (i.e. one that simply describes the shape of the SAD), which can then be used as a system descriptor in further analyses. In a recent paper, Ugland *et al.,* (2007) suggested that a mixed gamma binomial distribution ('gambin') meets these criteria. Gambin is a stochastic model that combines the gamma distribution with a binomial sampling method. The distribution has a single parameter (alpha: α), which describes the shape of the distribution.

A common context in which the form of SADs is of interest is in analyses of community responses to disturbance and land use change (e.g. Ugland & Gray, 1982; Mac Nally, 2007; Dornelas *et al.,* 2011). For such purposes, a model must provide a good fit to a variety of SAD shapes, and must possess a parameter which summarises, and tracks change in, the shape of the distribution and thus allows a comparison between the SADs of undisturbed and disturbed communities. The limited application of gambin thus far has

produced promising results, with good fits to empirical data (Ugland *et al.*, 2007). However, the lack of easily accessible methods for fitting the gambin distribution to empirical datasets, and the absence of maximum likelihood methods for estimating α has limited the number of applications of the model. Here, we address these issues by presenting a maximum likelihood derivation of the gambin distribution in conjunction with an R package (*gambin*). We then use these methods to undertake a comprehensive evaluation of gambin. We present a rigorous test of the fit of the gambin distribution using a variety of datasets. First, we use an extensive and well-specified arthropod dataset from the Azorean archipelago: a dataset large enough (over 90,000 individuals) to permit a statistically powerful comparison of alternative SAD models. The dataset includes samples from a range of land use types, which allows us to test the applicability of α as an ecological indicator. As the form of the SAD may change with the spatial grain of sampling and the number of individuals in a sample (e.g. Preston, 1948; McGill, 2011; Borda-de-Água *et al.*, 2012) a SAD model must perform well across a variety of spatial scales, and be relatively independent of sample size, to be of general utility. Thus, we also test the utility of α across a range of spatial grain sizes. In addition, we sourced ten species abundance datasets from the literature, representing a range of taxa and ecological contexts. We use these datasets in combination with the Azorean data to address the following questions:

- Does the gambin distribution provide an adequate fit to the data?
- Are there are any parts of the empirical distribution in which gambin does or does not provide a good fit (cf. Connolly & Dornelas, 2011)?
- How does the fit of the gambin distribution compare to three popular competitor SAD models (Poisson lognormal, logseries and zero-sum multinomial)?
- Is α sensitive to the proportion of individuals sampled from a community (cf. McGill, 2011; Locey & White, 2013)?

- Is the performance of gambin constant across sampling grain sizes (e.g. Šizling *et al.*, 2009; Borda-de-Água *et al.*, 2012)?
- Can α differentiate between community structures in different land use types (cf. Ugland *et al.*, 2007; Dornelas *et al.*, 2011)?

10.4 Methods

10.4.1 Gambin distribution

Here we present a brief overview of the model, noting that a mathematical description of the gambin distribution is provided by Ugland *et al.*, (2007). The basic idea underpinning gambin is that although the abundance of a species is determined by a mixture of many deterministic and stochastic factors, it is possible to obtain a good approximation to the observed species abundance curves by modelling the population sizes of the species in a community in two steps. First, we define the fitness of a species as the probability of achieving a large population size, and represent the frequencies of fitness values (i.e. the probability density) across species by a gamma distribution with scale parameter fixed at the value 1. The gamma distribution was chosen as the basis of the model as it is known to be a flexible distribution and this flexibility is preserved when the number of parameters is reduced by fixing the scale parameter. The scale parameter of gambin is set to 1 as the scaling of the distribution is achieved by fixing the max octave (see below). This is a pragmatic choice as it allows for all octaves to be fitted whilst simultaneously reducing the number of model parameters to one. Thus, the shape parameter (α), which determines the form of the distribution, is the only free parameter. A small α induces a distribution skewed to the left, i.e. a high density at small abundance values. This is what is observed in logseries-like distributions. A high α induces a distribution closer to normal on a log scale of abundances.

The second step is to link the gamma distribution (i.e. the fitness frequencies) to the actual abundance values. As empirical SADs are highly right skewed, we log transform the

number of individuals into 'octaves' (see Gray *et al.*, 2006a), and then estimate the number of species in each octave (Ugland *et al.*, 2007). While binning has been criticised for resulting in the loss of information, it is unlikely to bias parameter estimates (Connolly & Dornelas, 2011). Furthermore, as empirical abundance data often contain a significant amount of variation due to sampling effects, the added precision of models that are not based on binned data is questionable. We created abundance octaves by a simple log2 transform that doubles the number of abundance classes within each octave (a sensitivity analysis demonstrated the robustness of the results to this binning procedure; see below). Thus, octave 0 contains the number of species with 1 individual, octave 1 the number of species with 2 or 3 individuals, octave 2 the number of species with 4 to 7 individuals, and so forth (see Appendix 8.1 for a more detailed description). This binning procedure is method 3 of Gray *et al.* (2006a), and is listed by that study as the most appropriate binning method from several reviewed. The assignment of a species abundance into octave x is then regarded as the result of a binomial process with x trials (see Appendix 8.1 and Ugland *et al.*, 2007).

10.4.2 A description of the *gambin* R package

The full derivation of the maximum likelihood estimation of gambin, which is novel to this paper, is presented in Appendix 8.1 and is incorporated in an R package (*gambin*; version 1.0 accepted by CRAN in September 2013). Examples of use of the *gambin* package are also presented in Appendix 8.1. The *gambin* package contains functions to calculate the gambin distribution (**dgambin**) and to fit the gambin distribution to empirical data using maximum likelihood (**fitGambin**), along with a set of utility functions. The **dgambin** function returns the density function of the gambin distribution, given α and the octave number of the most abundant species. The syntax is similar to e.g. the basic density functions in the base package *stats* (e.g. **dnorm)**. To get the gambin distribution in units of species, the **gambin_exp** function is used.

The **fitGambin** function accepts a vector of abundances (optionally using a subsample of the individuals), which it bins into octaves using the utility function **create_octaves**. It then uses optimisation algorithms in R to identify the value of α that maximises the likelihood function. The return value of **fitGambin** is an S3 object of class 'gambin'. This class provides **print**, **confint**, **predict** and **plot** functions: **confint** gives the 95% confidence interval around the estimated α value; **predict** gives the predicted number of species in each octave; and **plot** creates a bar graph comparing empirical abundances (shown as grey bars) to the predicted values from the gambin distribution (shown as black dots).

10.4.3 Azorean arthropod data

Our main test dataset forms part of a long-term (1999–2012) biological study, the BALA ('**B**iodiversity of **A**rthropods from the **L**aurisilva of the **A**zores') project (see Borges *et al.*, 2005; Ribeiro *et al.*, 2005). Eighteen fragments of protected native Laurisilva forests were sampled for arthropods across seven islands using 100 randomly located transects (150m x 5m). Datasets were created by pooling the fragment samples present on each individual island. One island (Santa Maria) only has a single fragment and was thus not used as an island dataset (see Table A8.1 and A2 in Appendix 8.2 for site and sample details). Thus, we used 18 fragment samples and six island samples in this study. Along each transect, ground surface arthropods (largely epigean) were surveyed using 30 pitfall traps (Borges *et al.*, 2005), while canopy species were surveyed using a beating tray methodology (Ribeiro *et al.*, 2005) focused on three primary tree species. In all, 6770 samples (3420 pitfall traps and 3350 beating samples) were used for the current study (for further details see Gaspar *et al.*, 2008). In addition to the native Laurisilva fragments we used data from a related study on several of the islands, sampling epigean soil arthropods by the same pitfall methods for two additional land-uses – exotic plantation forest, and pasture (Cardoso *et al.*, 2009, 2013).

10.4.4 Model comparison

Both the 18 native forest fragment samples and the six island samples were used in our first model comparison, resulting in 24 different samples. Our first step was to fit the gambin distribution to each sample using the *gambin* R Package. We then fit the two most commonly used statistical distributions (PLN and logseries). To fit the PLN we used the *poilog* R package (Grøtan & Engen, 2009). The zero-sum multinomial distribution (ZSM) predicted by neutral theory was also fit to the 24 samples using the analytical form and likelihood function derived from Etienne (2005). We used two approaches to compare the goodness of fit of the various models. First, we compared the models using a Pearson's chi-square (χ^2) goodness of fit test. As the choice of goodness of fit test has been found to influence results (McGill, 2003) we also used a Kolmogorov-Smirnov test. Results were found to be the same using either test and so we present only the χ^2 test results herein. Sole reliance on traditional goodness of metrics such as χ^2 has been criticised as unreliable (e.g. McGill, 2003). Thus, we also used an information theoretic approach (Burnham & Anderson, 2002) to compare the fit of gambin to that of the three other models (i.e. the PLN, logseries and ZSM). Model performance was compared using the Bayesian Information Criterion (BIC) and Akaike's Information Criterion corrected for small sample size (AIC_c). The smallest BIC and AIC_c values were taken to represent the single best model for a given sample providing that ΔBIC or ΔAIC_c to the next best model was >2 (Burnham & Anderson, 2002). The PLN was considered a two parameter model, the ZSM a three parameter model, and gambin and the logseries one parameter models.

Unlike the other distributions in our multi-model comparison, the maximum likelihood estimate of the ZSM is very sensitive to the initial parameters used in the optimisation process. As such, multiple initial parameter values were used in the model fitting process along with different optimisation algorithms in R ('optim' & 'mle2'). As the

gambin distribution is calculated using octaves, the other distributions in the model selection framework were also fitted to binned data using the same binning process as described above. The ZSM analysis was restricted to the complete assemblage (i.e. not the sample grain subsets and additional datasets; see below) as the model fitting process was very time intensive. This omission is unlikely to affect our estimation of gambin's performance as initial observations found the ZSM only rarely provided a superior fit to gambin.

While statistical approaches to assessing goodness of fit are evidently more objective, visually analysing the fitted distribution can provide valuable information (Connolly & Dornelas, 2011). Thus, for each of the 24 samples we plotted the SAD with the fitted gambin model and visually inspected the fit to determine if there are particular parts of the SAD or types of SAD patterns where gambin does or does not provide a good fit.

10.4.5 Sourced datasets

We also compiled ten additional datasets widely used in SAD studies, representing a variety of systems and taxa (Table 10.1; see Appendix 8.2 for full acknowledgments to the data providers). Although a number of these datasets were used in the original gambin paper by Ugland *et al.* (2007), both our software and the determination of goodness of fit are novel to the present study and the inclusion of these datasets thus permits direct comparison with Ugland *et al.*, (2007) and with other comparative SAD analyses that have used these datasets. We again deployed the multi-model comparison for each dataset (excluding the ZSM); examining the fit using the aforementioned graphical methods at each stage.

10.4.6 Sample size and grain size

To determine the sensitivity of model parameters (gambin's α, the mean of the PLN, and the alpha of the logseries) to the proportion of individuals in a sample, we simulated a large number of different metacommunities of two types, lognormal (i.e. the SAD of the metacommunity was lognormal) and logseries. The set of lognormal metacommunities

comprised subsets of metacommunities with 50, 100, 500 and 1000 species. For each category of species richness, we simulated metacommunities with the number of individuals varying from 1000 to 1,000,000. There were 9 lognormal metacommunities simulated (see Table A8.3 in Appendix 8.3 for their exact properties). In addition, six logseries metacommunities were simulated, comprising subsets with 50, 100 and 500 species (each with 1000 and 10,000 individuals, details in Table A8.4 in Appendix 8.3). This range of metacommunities covered the two extremes of observed empirical distributions (i.e. lognormal and logseries) and a wide range in number of individuals (N) and number of species (S). For each of the described metacommunities we randomly sampled 1%, 10%, 25% and 50% of individuals using functionality in the *gambin* R package. We then fit the three distributions to each sample and to the full community (i.e. 100% of individuals), recording the parameters of each distribution in addition to the maximum and modal octaves (see Locey & White, 2013) of the observed distribution. This procedure was repeated 100 times in each case.

Table 10.1 Additional datasets used in the multi-model comparison and the associated dataset characteristics: location, taxon, number of species (S) and number of individuals (N). The full related acknowledgements are presented in Appendix 8.2. The best model was determined by comparing gambin with the Poisson lognormal distribution (PLN) and the logseries distribution using both a Pearson's χ^2 goodness of fit test and an information theoretic approach. PLN has two parameters and gambin and the logseries are single parameter models. A model was selected as best by means of the information theoretic approaches if it had the lowest BIC and AIC_c using a minimum difference in both criteria of two. + relates to cases where none of the distributions provided a satisfactory fit.

Dataset Location	Taxon	S	N	Source	Best Model	
					χ^2	AIC_c & BIC
Barro Colorado Island, Panama	Trees (>10cm)	229	20852	Hubbell *et al.*, (2005)	Gambin	Gambin
UK	Birds	503	151781104	Baker *et al.*, (2006)	+	+
Firth of Clyde, Scotland	Marine nematodes	113	8896	Lambshead (1986)	Gambin	Gambin
Australia	Corals	154	44225	Dornelas & Connolly (2008)	Gambin	Gambin
English Channel	Marine nematodes	321	1200	Unpublished data	Gambin	Gambin
Irish Sea	Marine nematodes	178	58372	Lambshead & Boucher (2003)	Gambin	Gambin
Rothamsted, UK	Lepidoptera	195	6813	Williams (1964)	Gambin	Gambin
Pasoh Forest Plot, Malaysia	Trees (>1cm)	808	295133	FRIM (2013)	Gambin	Gambin
Hinkley Point, UK	Fish (1981-2003)	82	3891	Magurran & Henderson (2003)	Gambin	Gambin
Sherman Forest Plot, Panama	Trees	129	3363	Condit (1998)	Gambin	Gambin

In addition, for a subset of four of the simulated lognormal metacommunities representing a range of S and N, 19 samples were taken, ranging from 5% of individuals in each metacommunity, to 95% of individuals, at intervals of 5%. The three models (gambin, PLN, and logseries) were fitted to the samples and the relevant parameters recorded. This process was repeated 100 times and the mean of the parameter values calculated. We then

calculated the sample size needed in order for the parameter estimates of the sample to be within 10%, 25% and 50% of the parameter estimates for the whole community.

To test whether the fit of gambin was consistent across sampling grain sizes, three native Azorean forest fragments that contained large numbers of transects were selected. For each of these fragments the multi-model comparison (excluding the ZSM) was calculated for a sample consisting of a single transect, then two transects and so on iteratively up to the maximum number of transects in a given fragment. Species abundances were averaged across all possible combinations of transects in each case. So, for example, when creating a sample from two transects, we averaged species abundances across all possible combinations of two transects within that fragment. All samples from each island were combined into island-level datasets as examples of larger grain sizes. It was not necessary to keep N constant to calculate α in this particular analysis as here we were testing whether the fit of gambin is consistent across a range of sample sizes (i.e. a range of N).

10.4.7 Comparison of alpha values between land use types

To evaluate the performance of gambin's α as an ecological indicator, we ranked the different habitat types in the Azorean dataset from untransformed (native forest) through moderately transformed (non-native forest plantations) to highly transformed (agricultural pasture). Three islands, Terceira, Flores and Faial, had suitable sample sizes for each habitat type and for these purposes each island was treated as a distinct system. Within each habitat type the gambin distribution was fitted to each transect and the α value recorded. This was done using a re-sampled set of samples using a fixed N value determined by the number of individuals in the least populated transect in order to remove any bias due to differences in sample size between land uses. A Wilcoxon rank sum test was used to determine whether α varied significantly as a function of land use. This method was repeated for the alpha parameter of

the logseries and the sigma parameter of the PLN, again after accounting for differences in sample size.

10.4.8 Testing the sensitivity of alpha to the binning method

While our binning method (log2) is frequently employed in SAD studies, the choice of log2 is somewhat arbitrary and theoretically any base could be used to bin the data into octaves. Thus, to test the robustness of the decision to use log2, we conducted two analyses. First, we calculated gambin's α for the 18 fragments using our standard log2 binning method and compared these values to α derived using a variety of other bases (base e, base 3, base 4 and base 5). In order to compare alpha values derived from these different binning methods we plotted pairwise comparisons of each α set and tested the degree of correlation in each instance using Pearson's product-moment correlation. Second, we undertook a simulation approach. For each run of the analysis, the aforementioned metacommunity simulation method was used to simulate communities with a) 10,000 individuals, and b) 100,000 individuals. For the first iteration of the run, the number of species was set to 50. At each subsequent iteration, species number was increased by 50, up to the maximum of 1000 species (i.e. 20 iterations). At each stage the gambin model was fit to the data using the five different binning methods, and the α values recorded. At the end of each run, a Pearson's product moment correlation was calculated for each pairwise comparison of α sets (i.e. different binning methods) and the values stored. The number of runs was set to 100 and the mean of the correlation values, along with the corresponding standard deviation, was calculated for each pairwise comparison. We were unable to test bases higher than 5 due to constraints on the number of individuals in our samples. For example, use of base ten resulted in only three octaves for certain samples. All analyses were conducted in R (R Development Core Team, 2012).

10.5 Results

10.5.1 Model comparison and inspection of the fit

For the Azorean fragment and island model comparison analyses, gambin provided a better fit to the data according to the Pearson's χ^2 test for all but six of the fragments (Table 10.2), and for all but one of the six islands (Table A8.5 in Appendix 8.3). For three fragments (16–18), the PLN provided a better fit, and for three fragments (7, 10 & 15) and one island (3: Pico) the ZSM provided a better fit, according to χ^2. Gambin outperformed the other three distributions according to both BIC and AIC_c for all 24 samples (Table 10.2 & Table A8.5). Gambin also performed best for all of the non-Azorean datasets according to χ^2 tests (Table 10.1, see Fig. A8.1 in Appendix 8.3 for examples), and for nine of the ten datasets according to AIC_c and BIC. None of the models provided an adequate fit for the British breeding bird data.

Visual examination suggested that gambin provides a good fit to a variety of empirically observed distribution shapes (e.g. Fig. 10.1a, see also Fig. A8.1 in Appendix 8.3), ranging from logseries-like to lognormal-like patterns. Equally, the fit is consistent across the different parts of the SAD, i.e., for both the rarer species and more common species. Nonetheless, visually the fit does not appear to be as good for those samples which exhibited a degree of multi-modality within the SAD (e.g. Fig. 10.1b).

10.5.2 The effect of sample size

Our analyses showed that when simulating lognormal metacommunities, the shape of the observed SAD, and thus the value of gambin's α, varied as a function of the proportion of individuals sampled (Tables A4 and A5 in Appendix 8.3). Fig. 10.2 illustrates this for one particular lognormal metacommunity. When basing analysis on a small sample, the observed SAD exhibits a logseries shape (Fig. 10.2a), but as the sample size increases, the shape of the SAD becomes more lognormal and the modal and maximum octaves both shift to the right

(Fig. 10.2c,d), as predicted by Preston's (1948) veil line concept. The mean and sigma of the PLN, and the alpha parameter of the logseries distribution, also change notably with the proportion of individuals in the sample (Fig. 10.2). For the logseries metacommunities, the change in gambin's α with proportion of individuals sampled was not as pronounced as for the lognormal metacommunities.

Gambin and the PLN generally required similar sample sizes in order for the parameter estimates of the sample to asymptote towards the parameter estimates of the whole community; the logseries generally required slightly smaller sample sizes (see Table A8.6 in Appendix 8.3 for the full results). For example, for a sample estimate to be within 10% of the whole metacommunity parameter estimate, the proportion of individuals sampled from the metacommunity needed to be on average: 68% (range: 55–90%) for gambin's α, 65% (range: 60–70%) for the mean of the PLN, and 53% for the logseries' alpha (range: 45%–60%; see Table A8.6 for the equivalent results based on 25% and 50% accuracy).

The performance of gambin was consistent across sample grain sizes. For each of the three Azorean fragments (total of 24 samples) used in the sample size analysis, gambin outperformed the PLN and logseries distribution for each combination of transect number according to BIC and AIC_c (see Table A8.7 in Appendix 8.3 for results for all three fragments).

Table 10.2 Goodness of fit and model selection results for arthropod SADs of 18 (No.) native Laurisilva forest fragments in the Azores. Arthropods were sampled using a standardised pitfall trap and canopy beating methodology between 1999 and 2004. For each fragment the Pearson's c^2 statistic and associated p value (in parentheses) are presented for the gambin, logseries, and Poisson lognormal (PLN). The Bayesian information criterion (BIC) and Akaike's information criterion corrected for small sample size (AIC_c) are also given for all four distributions. PLN has two parameters and the ZSM has three parameters. Gambin and the logseries are single parameter models. The best model according to each criterion, using a minimum difference of two, is highlighted in bold for each fragment. Fragment information can be found in Supplementary material Appendix 8.2, Table A8.1

	Gambin			Logseries			PLN		
No.	χ^2 (*P*)	BIC	AICc	χ^2 (p)	BIC	AIC_c	χ^2 (*P*)	BIC	AIC_c
1	**19.6 (0.03)**	**395.6**	**395.7**	43.1 (0.00)	421.3	421.3	24.0 (0.01)	407.7	408.4
2	**3.8 (0.92)**	**339.0**	**339.2**	31.9 (0.00)	377.1	377.3	6.6 (0.68)	347.2	348.3
3	**11.3 (0.34)**	**491.2**	**491.3**	27.1 (0.00)	515.4	515.4	16.8 (0.08)	505.5	506.2
4	**8.2 (0.51)**	**365.9**	**366.1**	28.4 (0.00)	393.6	393.8	12.1 (0.21)	375.4	376.5
5	**10.9 (0.36)**	**413.1**	**413.1**	28.9 (0.00)	439.3	439.4	15.0 (0.13)	424.4	425.11
6	**23.7 (0.01)**	**412.1**	**412.2**	54.8 (0.00)	446.9	447.0	28.0 (0.00)	424.2	424.9
7	25.5 (0.00)	**526.5**	**526.5**	50.7 (0.00)	563.4	563.5	25.5(0.00)	539.1	539.8
8	**19.0 (0.06**)	**458.4**	**458.3**	36.6 (0.00)	480.4	480.3	26.9 (0.00)	474.1	474.5
9	**9.2 (0.51)**	**418.3**	**418.3**	27.8 (0.00)	445.8	445.8	13.2 (0.21)	427.1	427.8
10	19.5 (0.01)	**531.8**	**532.2**	30.9 (0.00)	559.2	559.5	18.7 (0.02)	544.2	545.8
11	**7.5 (0.67)**	**331.2**	**331.2**	19.8 (0.03)	346.8	346.9	11.9 (0.29)	340.1	340.8
12	23.6 (0.00)	**510.7**	**510.9**	37.3 (0.00)	538.7	538.9	**23.2 (0.01)**	523.7	524.8
13	**25.2 (0.00**)	**582.9**	**582.9**	47.1 (0.00)	615.0	615.0	32.8(0.00)	596.6	597.3
14	**15.7 (0.07)**	**485.7**	**485.9**	33.0 (0.00)	514.2	514.4	18.3 (0.03)	497.5	498.6
15	24.9 (0.00)	**387.2**	**387.4**	26.2 (0.00)	404.4	404.6	17.0 (0.05)	394.8	395.9
16	28.6 (0.00)	**467.1**	**467.5**	31.1 (0.00)	487.1	487.4	**22.2 (0.00)**	476.6	478.2
17	30.3 (0.00)	**244.4**	**245.2**	46.2 (0.00)	261.9	262.8	**27.0 (0.00)**	250.4	253.5
18	29.9 (0.00)	**440.2**	**440.6**	37.5 (0.00)	460.5	460.9	**24.8 (0.00)**	449.5	451.1

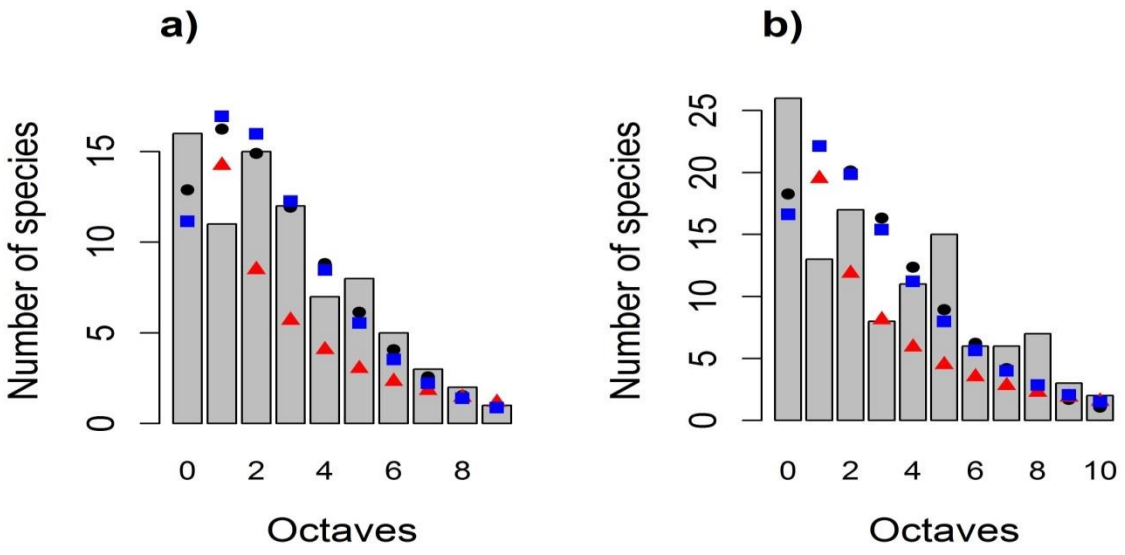


Figure 10.1 Examples of the fit of the gambin distribution (black dots), the logseries (red triangles), and Poisson lognormal (blue squares) to observed data (bars). For reasons of clarity, the predicted values of the logseries have occasionally been omitted when they are much larger than the observed values. (a) Data are for arthropods from a fragment (No. 2, Supplementary material Appendix 8.2, Table A8.1) of native Laurisilva forest on the island of Faial, Azores. The a parameter of the gambin distribution is 1.9. (b) Data are for arthropods from a fragment (7, Supplementary material Appendix 8.2, Table A8.1) of native Laurisilva forest on the island of Pico, Azores. The α parameter of the gambin distribution is 1.8. In both plots, gambin provides the best fit according to both BIC and AIC_c.

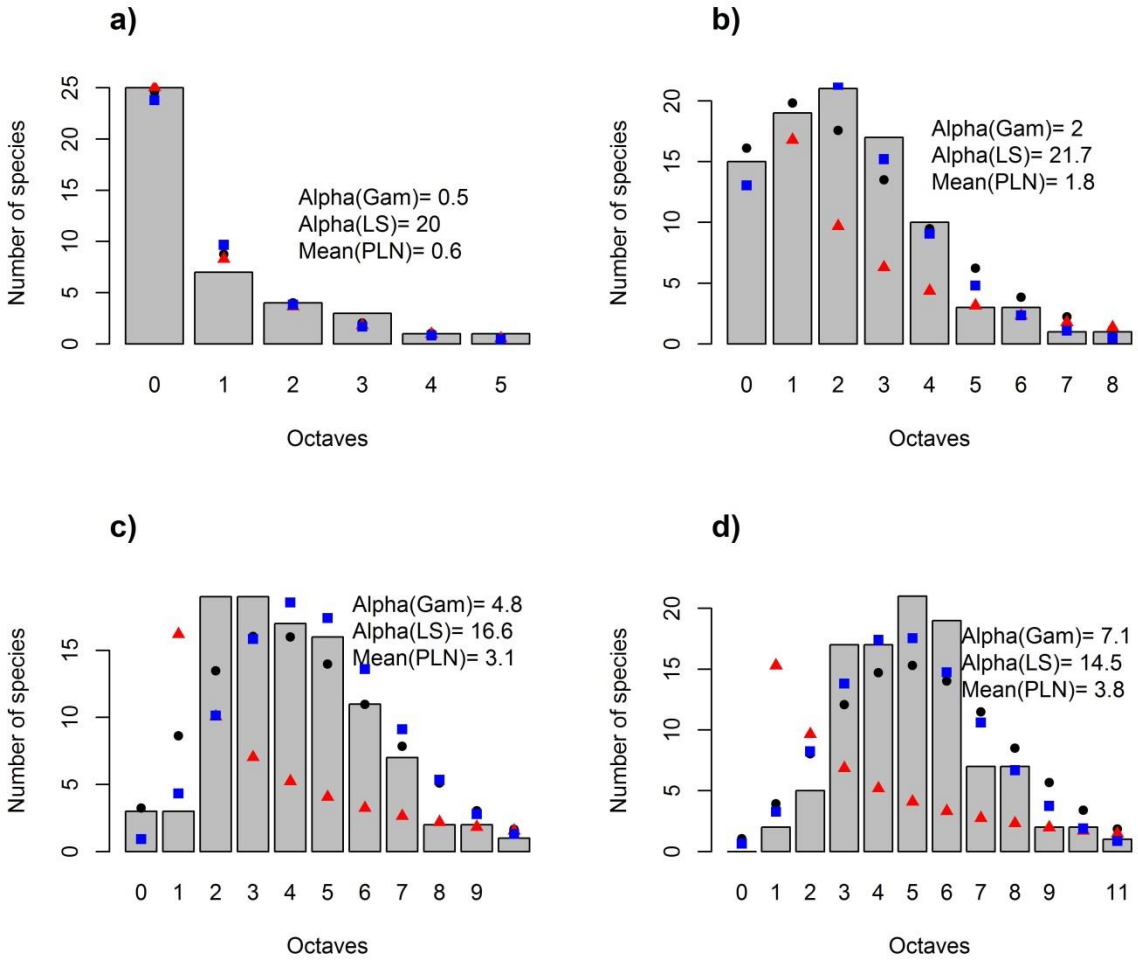


Figure 10.2 Changing shape of the observed species abundance distribution with sample size. Data were sampled from a simulated lognormal metacommunity (number of individuals (N) = 10 000, number of species = 100). The four plots correspond to different levels of sampling from the metacommunity: (a) 1% of individuals (N = 100) sampled from the metacommunity, (b) 10% of individuals (1000), (c) 50% of individuals (5000), and (d) 100% of individuals, i.e. the full metacommunity (N = 10 000). On each plot the a parameter of the gambin distribution (Gam), the alpha of the logseries distribution (LS), and the mean of the Poisson lognormal distribution (PLN) calculated using the sampled data are given. The fit of the gambin distribution (black dots), PLN (blue squares) and logseries (red triangles) to each sample is also presented. For reasons of clarity, the predicted values of the logseries have been omitted from bin 0 when they are much larger than the observed values

10.5.3 Land use gradient

Considering Terceira Island, gambin's α was found to differ significantly between habitat types, with the highest mean α belonging to the native forest samples, and the lowest mean α belonging to the pasture samples (Fig. 10.3a). The results were similar for Faial and Flores, except that for these islands the differences between native forest and exotic plantation forest

were not significant at the 0.05 level (Table 10.3; see also Fig. 10.3b,c). The alpha of the logseries significantly differentiated between communities in the different land uses (i.e. $P < 0.05$) in two instances (native forest and pasture on Terceira, and native and exotic forest on Flores; see Fig. A8.2 in Appendix 8.3). The sigma of the PLN did not significantly differentiate between any of the communities (Fig. A8.2).

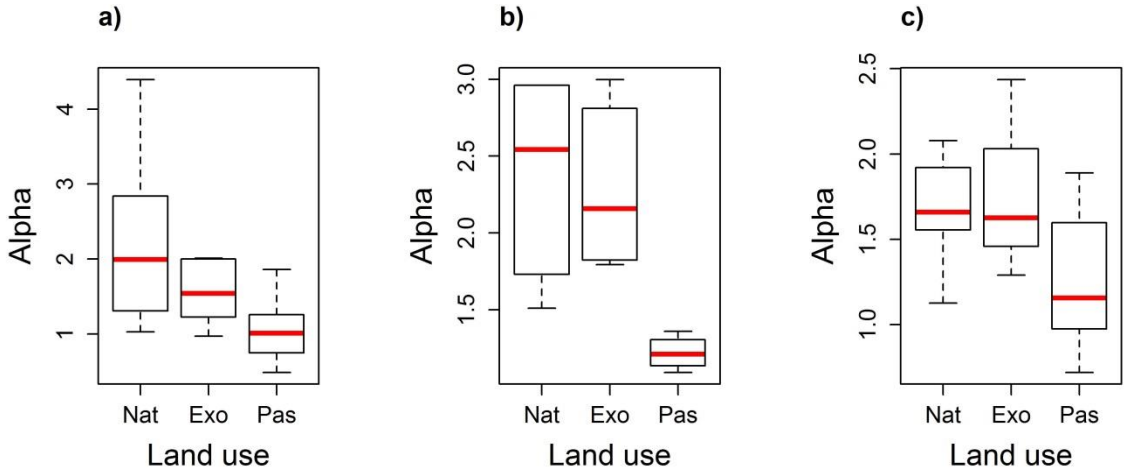


Figure 10.3 Difference in the gambin distribution α parameter values (standardised to keep sample size constant) between transects of the three land use types (native forest, exotic plantation forest, and pasture) on three Azorean islands: a) Terceira, b) Faial, c) Flores. The number of individuals used to standardise the calculations in each instance is given in Table A8.2 in Appendix 8.2. The box plots display the median (red line), the first and third quartiles (black box), and the minimum and maximum values (whiskers). The significance of differences between land use types according to Wilcoxon rank sum tests are presented in Table 10.3.

10.5.4 Sensitivity to the binning method

The use of log2 to bin the data into octaves appears to be a robust choice. For the 18 Azorean fragments the pairwise comparisons of the alpha values revealed them all to be highly significantly positively correlated ($P < 0.001$ in each instance, Fig. A8.3 in Appendix 8.3).

The comparisons were also highly correlated according to the simulations (Table A8.8 in Appendix 8.3).

Table 10.3 Wilcoxon rank sum test results for the significance of difference of α values of the gambin distribution between transects of different land use groups, for three Azorean islands: Faial, Flores and Terceira. For each island the analysis was run using the standardised re-sampled island samples (i.e. all transects were re-sampled in order to keep sample size constant). The number of individuals used to standardise the calculations in each instance is given in Table A8.2 in Appendix 8.2. Significance was set at the 0.05 level and all significant *P* values are highlighted in bold. The three land-uses were native Laurisilva forest, exotic plantation forest and pasture. The sample size (i.e. number of transects) for each grouping was as follows. Faial: native forest (8), exotic forest (6), and pastures (8). Flores: native forest (12), exotic forest (4), and pastures (8). Terceira: native forest (33), exotic forest (13), and pastures (35).

Island	Test statistics	Native/Exotic	Native/Pastures	Exotic/Pastures
Faial	W	22	15	1
	P value	0.53	**0.036**	**0.016**
Flores	W	12	32	29
	P value	0.864	**0.034**	**0.028**
Terceira	W	195	1210	80
	P value	**0.05**	**<0.001**	**0.001**

10.6 Discussion

We derived a maximum likelihood estimation of the gambin distribution and used it to undertake a comprehensive test of the model. We found that gambin is a flexible model, which provides a superior fit to empirical SADs compared to three popular alternatives. This flexibility, with only a single free parameter, is advantageous as it enables gambin to provide a good and parsimonious fit to a variety of empirical data. Also, the α parameter, which summarises the shape of the model, is of analytical utility as it can potentially be used, for example, as an explanatory variable in regression models exploring how environmental properties influence the SAD. Multi-parameter models, such as the PLN and ZSM, do not have this capability, and while the logseries also possesses a single parameter, the logseries is not a very flexible distribution and often provides a poor fit to empirical data (see e.g. Table 10.2). The flexibility of gambin is also beneficial because the shape of empirical SADs

change as a result of many processes; for instance, as a function of spatial grain size and as the result of ecological disturbance. Gambin is able to track these changes in SAD shape, and as this information is captured in the single parameter, α can be used to compare community SADs analytically.

10.6.1 Goodness of fit and multi-model comparison

Despite only having a single shape parameter (α), gambin provided a better fit than the other models in the majority of instances, including both Azorean and non-Azorean datasets. This included distribution patterns where the modal class was the singleton octave (i.e. logseries form), and patterns where the modal class represented an octave of more common species (i.e. lognormal form). Additionally, the fit provided a more accurate representation of the data than the other distributions for the various sample grain subsets (Table A8.7). It can be highly informative to examine how fit changes with scale. Our analyses show that, relative to the other distributions, both the fit and performance of gambin remained consistently good at spatial scales ranging from a single transect up to fragment- and island- scales of analysis. The different scales represent contrasting points along a continuum of community structure; the good fit of gambin to all subsets illustrates the potential of the model for SAD research.

The gambin distribution is restricted to be unimodal, and thus the poor fit to those SADs that visually exhibit multimodality is unsurprising (e.g. Fig. 10.1). Recent work has suggested multimodal SADs may be more prevalent than previously assumed (Fig. 10.1b, Dornelas & Connolly, 2008; Vergnon *et al.,* 2012; Matthews *et al.,* 2014c). However, functions capable of generating multimodal distributions tend to be complex. For instance, the tri-modal Poisson lognormal (PLN3) distribution (Dornelas & Connolly 2008; Matthews *et al.,* 2014c) possesses eight parameters compared to gambin's one. Moreover, complex distributions such as the ZSM and PLN3 can be problematic to fit, with an increased chance of locating local maxima when deriving the log-likelihood. While gambin cannot capture

multimodality, to the extent that distributions analysed herein are in fact multimodal in character, the gambin model nonetheless provided a better statistical fit than the logseries, PLN, and the ZSM.

10.6.2 Model performance and sample size

That the shape of the SAD is not a constant property of any given community is well known (Magurran, 2007). However, the separate issue of sample SADs being poor representations of community SADs when sample size is low is often overlooked (but see Preston, 1948; Pielou, 1975; Green & Plotkin, 2007; McGill, 2011; Locey & White, 2013). Our simulations indicate that this issue is particularly problematic when the SAD of the community being sampled is lognormal, bringing into question the utility of SADs when sample sizes are low. In particular, it appears that if the community SAD is lognormal, a sample size of 10% of the community's individuals will result in a sample parameter estimate that may differ from the community α estimate by up to 50%. This issue arises because the shape of the sample SAD changes considerably with varying sample size in relation to the community being sampled (Fig. 10.2), and as outlined above, the flexibility of gambin means the model accurately tracks this change in SAD shape. This small sample problem appears to be general, as it affects the parameters of the PLN and logseries distributions as well as gambin (Table A8.6; see also Pielou, 1975; McGill, 2011). The relative stability in gambin's α for samples from the logseries metacommunities makes intuitive sense, as the modal octave of each sample always corresponded to the singletons' octave. Providing specific recommendations regarding a minimum value of N necessary to obtain accurate estimates of gambin's α is difficult as the accuracy depends in part on the size of the community being sampled, which in empirical systems is often impractical to calculate. However, as an indicative guideline, we agree with McGill (2011), who suggests a minimum N of 1000 individuals.

As gambin's α varies with sample size, any rigorous comparative analysis of α should be based on keeping N constant across samples. A recent study by Locey & White (2013) drew similar conclusions. Focusing on SADs and the feasible set concept, they reported that the specific form of the SAD is constrained by N, and that differences in the form of the SAD between communities may result purely from differences in sample size. Our methodology of comparing gambin α values using repeated re-sampling to a common N (using functionality provided in the gambin R package) provides a way of circumventing the sample size problem, and our comparative analysis of variations in α between land use types (Table 3) indicates that gambin α can have discriminatory power as an ecological indicator when used in this way. We advocate this method more generally in SAD research, as our analyses indicate the parameters of various SAD models (i.e. not just gambin) are related to the number of individuals in the sample and thus using the re-sampling method provides an unbiased way of using model parameters in comparative analysis.

We are also confident that our choice of log2 to bin the data into octaves is appropriate. While using different bases did result in small differences between alpha values, each set of alpha values was highly correlated with every other set (Fig. A8.2 and Table A8.8 in Appendix 8.3). Thus, the behaviour of alpha was consistent irrespective of the binning method employed and we are confident our study findings are not dependent on the log base used.

10.6.3 The alpha parameter as an ecological metric

It has long been appreciated that the SAD of undisturbed communities (i.e. those in equilibrium) resemble the lognormal, while disturbed communities more closely follow a logseries SAD (Gray & Mirza, 1979' Ugland & Gray, 1982). We have shown that the gambin distribution is flexible, meaning it can fit a variety of empirical SAD shapes (including lognormal and logseries -like shapes), and that the distribution shape is adequately

characterised by the model's single parameter (α): low values of alpha indicate logseries-like SADs, and high alpha values indicate lognormal-like SADs. Again, the only other distribution with a single parameter which can be used in a similar way is the logseries, and this is much less flexible than gambin and fits far fewer datasets.

Our Azorean analyses support the claim (Ugland *et al.*, 2007) that gambin's α is a useful metric with which to compare communities affected by different types of disturbance or habitat transformation, as they showed that α generally decreased from the relatively untransformed native forest transects, to exotic plantation forest, to the highly transformed pastures (cf. Meijer *et al.*, 2011; Cardoso *et al.*, 2013). These differences were significant in seven out of nine cases, indicating that this metric provides a reasonably sensitive diagnostic tool. Land use change is frequently cited as the main driver of current biodiversity loss (e.g. Fischer & Lindenmayer, 2007), and thus any ecological metric which can effectively characterise the impact of land use change on the SAD, and community structure more generally, is a useful addition to the ecologist's statistical toolbox. Furthermore, this finding is of particular interest considering that the logseries and PLN distributions were generally unable to differentiate between the different land uses to the same degree (Fig. A8.3).

That α fails to differentiate between Azorean native and exotic forest in Faial and Flores simply indicates that the SADs of these two land uses are similar in form, for these islands. This is an ecologically interesting result, which likely reflects the large number of introduced arthropod species on the Azores. The majority of these introduced species are well adapted to exotic forest habitats in the archipelago, and are often sampled in relatively high abundances (Cardoso *et al.*, 2013). Thus, it is likely that the inclusion of such species obscures the impact of the loss of native Laurisilva forest on community structure, resulting in similar SAD shapes between forest types. However, removing the introduced species from

the analysis is not an option in this case as it often results in sample sizes too low to permit accurate estimation of α (above; see also McGill, 2011).

The lack of easily accessible methods to fit the gambin distribution hitherto, coupled with the absence of a maximum likelihood derivation, has restricted gambin's dissemination among the ecological community. The new methods and R package presented in this paper alleviate this issue and allow for both easy computation of α, and the incorporation of gambin within information theoretic model comparisons. Characterised by a single parameter that is analytically practical, can be easily interpreted, and provides flexibility, and taking into account the results of our tests, gambin presents a promising tool for future SAD research.

CHAPTER 11

Multimodal species abundance distributions: a deconstruction approach reveals the processes behind the pattern

Paper published in Oikos: Matthews, T.J., Borges, P.A.V. & Whittaker R.J. (2014) Multimodal species abundance distributions: a deconstruction approach reveals the processes behind the pattern. *Oikos*, **123**, 533-544.

11.1 Preface

This chapter is the eight empirical chapter, and assesses the prevalence of multimodal species abundance distributions (SADs) in Azorean arthropod assemblages, alongside an examination of the possible drivers of the pattern. Bi- and tri-modal SADs were observed for a number of Azorean forest fragments and islands. It was found that that this pattern was driven primarily by the inclusion of satellite, invasive and tourist species in the samples which created a mode of rare species in the SAD.

Authors: Thomas J. Matthews, Paulo A.V. Borges & Robert J. Whittaker

The co-author contributions were as follows. T.J.M. conceived the ideas; T.J.M. ran the analyses; and T.J.M. and R.J.W. wrote the paper with a significant contribution from P.A.V.B.

11.2 Abstract

It is becoming increasingly apparent that the species abundance distribution of many ecological communities contains multiple modes, a phenomenon that has been largely overlooked. Here, we test for multiple modes in the species abundance distribution using a combination of one, two and three mode Poisson lognormal distributions and an extensive arthropod dataset from the Azores. We consider the abundance distribution of twelve native Laurisilva forest fragments and the combination of fragments within five islands, allowing us to detect whether patterns are consistent across scales. An information theoretic approach is employed to determine the best model in each case. To explore the processes driving multimodal abundance distributions we tested various potential mechanisms. We classified species as core if they are present in over half the fragments in our study, and as satellite if they are sampled in fewer than half the fragments. Furthermore, species are classified based on body size, whether they are indigenous (i.e. endemic or native non-endemic) or introduced to the Azores, abundance in land uses other than native forest, and dispersal ability. We find that models incorporating multiple modes perform best for most fragments and islands. A large number of communities are bimodal, comprising a mode of very rare species and a mode of relatively common species. Deconstructing the full assemblages into their constituent subsets reveals that the combination of ecologically different groups of species into a single sample underpins the multimodal pattern. Specifically, the rarer mode prevailingly contains a higher proportion of satellite taxa, introduced species and species that are more adapted to anthropogenic land uses that surround the native forest.

11.3 Introduction

The observation that ecological systems contain a small number of very abundant species and numerous relatively rare species is often described as one of ecology's few universal laws, with very few known deviations (Fisher *et al.,* 1943; Preston, 1948; Pielou, 1969; McGill *et*

al., 2007; Ulrich *et al.,* 2010). A species abundance distribution (SAD) describes the abundance of all species recorded within an ecological community of interest (Ulrich *et al.,* 2010). Over 25 different functions have been proposed to model SADs (McGill *et al.,* 2007; McGill, 2011); the two most widely used being the logseries (Fisher *et al.,* 1943) and the lognormal (Preston, 1948). As the continuous lognormal distribution allows fractional abundances, the Poisson lognormal (Bulmer, 1974) is often used instead of the standard continuous lognormal. Several authors (e.g. McGill *et al.,* 2007) have called for a shift away from simply determining how well particular distributions fit different datasets, to a more process and mechanistic oriented approach. One particular emerging area of interest in this regard has been the increased recognition of the possibility of multiple modes within the SAD.

While the possibility of multimodal SADs has long been discussed (e.g. Pielou, 1969, Ugland & Gray, 1982; Gray *et al.,* 2005; Borges *et al.,* 2008; Borda-de-Água *et al.,* 2012) it has largely been overlooked and few studies explicitly test for multiple modes when fitting SAD models (but see Dornelas & Connolly; 2008, Vergnon *et al.,* 2012). Even when multiple modes have been observed, the pattern has often been dismissed as an uncommon anomaly (e.g. Gaston, 1994). Pielou (1969) was one of the earliest to acknowledge the possibility of multimodal SADs, postulating that increasing a sample by broadening the type of species included, for example from a focus on warblers to all bird species, could result in the combination of multiple logseries distributions, each with its own set of parameters. Ugland & Gray (1982) also debated multimodal distributions, this time in the context of communities disturbed away from equilibrium. They argued that a realistic community model should take account of the asymmetry in abundances typical of ecological communities, illustrating this idea by subdividing data for a community into three distinct abundance groups, roughly translated as rare, intermediately abundant and common (Ugland & Gray, 1982, see an

application in Borges *et al.*, 2008). Within each group the SAD is symmetric, which leads to a mixture of three lognormal distributions; the combination of which results in an asymmetric SAD for the whole assemblage. More recently, Gray *et al.* (2005, 2006b) found strong evidence of bimodal lognormal distributions for a mixture of marine and terrestrial datasets.

Dornelas and Connolly (2008) advanced this approach to focus on multimodal Poisson lognormal (herein 'PLN') distributions. They fitted a combination of one to four mode PLN distributions in addition to the logseries for a large sample of Australian coral communities and compared the various distributions using maximum likelihood methods. At the scale of the entire sample the three-mode PLN was found to have the greatest support (Dornelas & Connolly, 2008). Thus, it appears likely that multimodal SADs are common in nature; it is simply that they have been overlooked. Vergnon *et al.* (2012) provided further evidence supporting this conjecture. Sourcing a number of previously published datasets from the literature relating to an array of taxa, Vergnon *et al.* (2012) confirmed that multimodality is not simply confined to a few anomalous systems.

Numerous potential mechanisms driving the aforementioned multimodality have been proposed. The combination of multiple samples through space and/or time has been shown to result in the amalgamation (the 'amalgamation hypothesis') of two different groups of species (Hanski, 1982; Magurran & Henderson, 2003; Ulrich & Zalewski, 2006; Unterseher *et al.*, 2011). Firstly, there is a group of core species: those which are widely distributed and present/abundant within numerous samples (Ulrich & Zalewski, 2006). The second group, often termed satellite or occasional species (herein satellite), are only recorded in a few samples and are generally viewed as representing migrants from the surrounding species pool (Unterseher *et al.*, 2011). The two sets of species leave different signals on the SAD and when grouped together can result in the multimodal SADs which have been observed (Gray *et al.*, 2005). In addition to the core/satellite dichotomy, it is possible that the grouping of

other distinct species subsets into one sample may lead to multiple modes. For instance, dispersal ability has been postulated as an important determinant of a species' abundance within a community (Borda-de-Água *et al.,* 2007; Alonso *et al.,* 2008). It has also been suggested that introduced species may influence the SAD of a given community (Labra *et al.,* 2005): a process which is likely to be prevalent on many islands, as such systems have received a disproportionate amount of species introductions (Whittaker & Fernández-Palacios, 2007). Moreover, it is likely that these various subset divisions are not entirely independent; for example, it is possible that in general satellite species include many better dispersers than the core group and are present in the native forest due to mass effects (Shmida & Wilson, 1985). Deconstructing entire assemblages into these different subsets and analysing each subset for particular patterns individually has been advocated as an effective method for elucidating new perspectives on macroecological and biogeographical patterns (Marquet *et al.,* 2004).

Emergent neutrality theory (Scheffer & van Nes, 2006; Segura *et al.,* 2011; Vergnon *et al.,* 2012; herein 'ENT') has also been put forward as a potential explanation for multimodal SADs (Vergnon *et al.,* 2012). ENT contends that species are organised along a niche axis such that multiple modes of coexisting species are observable. The species within the 'core' of these niche axis modes are relatively abundant while the species in the 'valleys' are relatively rare, which results in a bimodal SAD (Vergnon *et al.,* 2012). Multimodal SADs are also observed if the species within the different niche axis modes differ significantly in terms of their abundance. Finally, neutral theory proper has also been shown to result in multimodal SADs (Borda-de-Água *et al.,* 2007), but testing this proposition is beyond the scope of the present study.

In this paper, we test for multimodal PLN SADs using an extensive arthropod dataset from the Azores comprising 12 fragments of native Laurisilva forest, across five islands.

Analyses are performed at both the fragment and island scale to determine whether the presence of multiple modes is dependent on the scale of amalgamation of samples. We will show that, while not a universal pattern, multiple modes are present in the SADs of numerous samples. In order to infer the processes driving the observed multimodal SADs, we test various potential explanatory variables. The full assemblage is deconstructed into various subsets, including core and satellite groups, and the proportions of these different subsets within the different modes examined. We also test the possibility that self-limiting similarity along a body size niche axis appropriately characterises Azorean Laurisilva forest arthropod assemblages.

11.4 Methods

11.4.1 Study site and sampling methods

The data used in the current manuscript are part of a long-term ecological study conducted in the Azores archipelago, the BALA project (1999-2012; **B**iodiversity of **A**rthropods from the **L**aurisilva of the **A**zores; see Borges *et al.*, 2005; Ribeiro *et al.*, 2005). Eighteen fragments of Laurisilva forest were sampled across seven Azorean islands for arthropod species using a combination of pitfall traps and beating trays. A more comprehensive summary of the BALA dataset and associated sampling methods is provided in Chapter * (section **). As fitting the various multimodal SAD models was very time intensive a subset of 12 fragments and five islands were selected from the BALA dataset for analysis (see Fig. 11.1). For each fragment the individual transects were grouped together to form the dataset for that particular fragment. The minimum number of transects sampled within the fragments is four, and thus in order to maintain consistency across the samples four transects were pooled in all fragments. For fragments in which more than four transects were sampled, four were selected at random. To create the island datasets the same transects that were combined to form the individual fragment samples were pooled for all fragments on a particular island. A detailed description

of the sampling methods and study sites can be found in Gaspar *et al.* (2008). Hereafter 'sample' refers to the unit of analysis for which a SAD was constructed, i.e. fragment or island. Thus, a reference to 'all samples' relates to all 12 fragment and five island datasets.

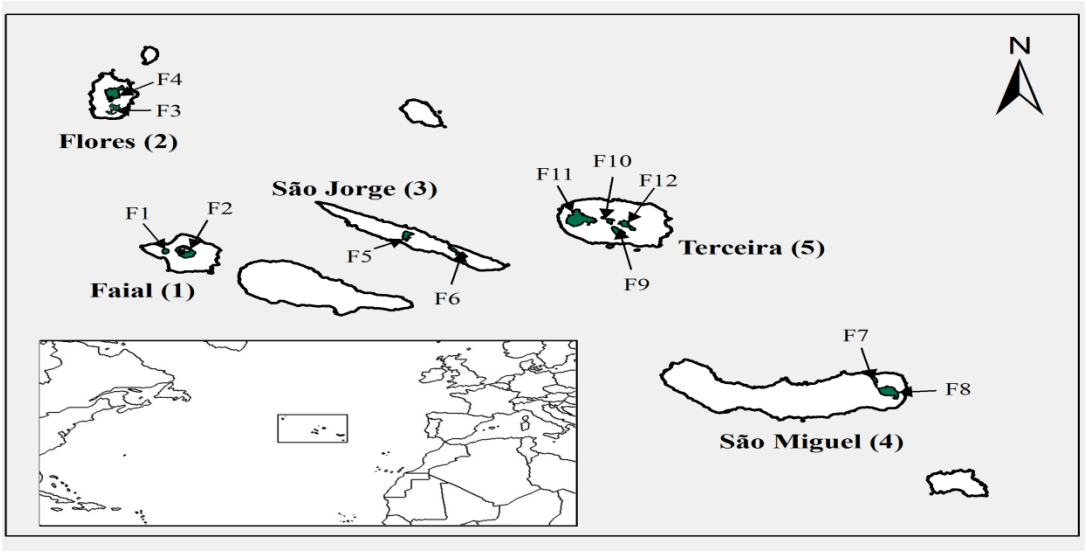


Figure 11.1 Map of study sites in the Azores. The names of the five islands are given and the numbers in parentheses refer to the corresponding island information in Table 11.2. Within these five islands the locations of the 12 native forest fragments are presented alongside the fragment codes which correspond to the individual fragment information in Table 11.1. Due to issues with clarity the figure is not to scale; in particular, the native forest fragments have been enlarged to allow them to appear visible.

11.4.2 Species classification

In order to facilitate greater understanding of the mechanisms behind multimodal SADs, the entire assemblage was deconstructed into distinct subsets. Species were classified as native or introduced, and as possessing either high or low dispersal ability by a trained Azorean taxonomist (PAVB). Following Magurran and Henderson (2003; see also Hanski 1982), species were further classified into core and satellite subgroups. However, this was achieved through focus on the pooling of species through space as opposed to time (see Gray *et al.*, 2005; Ulrich & Zalewski, 2006). A species was classified as a core species if it was present in more than 50% of the fragments in our study (cf. Ulrich & Zalewski, 2006), or a satellite species if it was present in 50% or less. However, as Ulrich & Ollik (2004) note, it is likely that there is a 'continuum' between core and satellite species and any dichotomy is likely to

be a simplification of the ecological reality. Therefore, by way of a sensitivity analysis, we also used a different core/satellite classification method to determine whether our choice of 50% affected results. For this purpose we split species into three groups: those present in 25% or fewer fragments, 26-75%, and 76-100% of fragments.

As part of a separate project three other land-uses in addition to native forest were sampled (see Cardoso *et al.*, 2009, 2013): exotic plantation forest, semi-natural pasture, and intensive pasture. This allowed a further classification of species based on the abundance of species in these other land uses. As such, species were classified as tourist species (see Borges *et al.*, 2008) if they were recorded as having a greater abundance in a land-use other than native forest. Thus, tourist species are either those species that are more adapted to anthropogenic land uses and are present in the native forest due to 'mass effect' like processes, or species which are habitat generalists and are found in large numbers in all four land uses. In order to keep sample size constant across land uses, transects were selected at random, and pooled, for each of the three additional land-uses based on the number of transects used to construct individual fragments. The body size of each species was also measured: i) for spiders, body length was collated from the literature, separately for males and females to address the possible effects of sexual dimorphism. As female and male body lengths were highly correlated (Pearson correlation=0.96; $p<0.001$), the average value between females and males was used; ii) for the other arthropods, we measured specimens available in our collection. Finally, species were classified as indigenous or introduced (above).

11.4.3 Statistical analyses

To determine whether the SADs of the various samples exhibited multiple modes, a combination of one, two and three mode Poisson Lognormal distributions (PLN, PLN2 and PLN3 respectively) were compared using a multi-model information theoretic framework

(Burnham & Anderson, 2002). The models were fit to each of the 12 fragments and five islands and model performance was calculated with the Bayesian Information Criterion (BIC) and Akaike's Information Criterion corrected for small sample size (AIC_c). The BIC is a useful measure in studies of this nature as it penalises to a greater extent for additional parameters than does AIC_c (Burnham & Anderson, 2004). As the PLN2 and PLN3 models have five and eight parameters respectively, the number of parameters is of particular importance in comparing model fits in our study. The AIC_c was preferred to the standard Akaike's Information Criterion (AIC) as it is more applicable when datasets are small, and because AIC_c converges to AIC when the sample size is large (Burnham & Anderson, 2004). The best model(s) within each sample was selected by setting a minimum difference criterion in BIC and AIC_c of two (Burnham & Anderson, 2002).

Due to the complexity involved in fitting the PLN2 and PLN3 distributions, and the increased chance of numerous local maxima, parameter searches were launched from multiple starting values, and the parameter values corresponding to the lowest negative log likelihood selected for use in the multi-model comparison. Initially, starting values were estimated from analysing the frequency distributions of the raw data. A more objective parameter search method was also conducted. This involved setting upper and lower limit boundaries for each parameter and then creating a loop that randomly selected values within these bounds for each parameter and then running the optimisation algorithm using these values, reporting the likelihood, parameter values and whether convergence was successful, after each iteration. Furthermore, a variety of maximum likelihood optimisation algorithms in R (i.e. 'optim', 'nlminb') were employed in the parameter searches.

The functions used to fit the PLN models were derived from Dornelas & Connolly (2008), particularly their supplementary material, and readers are directed there for the probability distribution formulae and full R code. In order to determine whether the PLN was

in fact a good choice of model to begin with, a logseries distribution was also fitted to each sample. A likelihood ratio test (LRT) based on a chi-squared distribution was also used following the multi-model comparison to determine if the increase in fit of more complex models was significant. Considering only nested models, a more complex model (i.e. more parameters) will always provide a better fit than a less complex model, and while AIC_c and BIC take the number of parameters into account, it can be useful to check the LRT statistic. The degrees of freedom for each model were the number of free parameters pertaining to each model. As LRT can only be used for nested models, the logseries was not included in this particular analysis. All analyses were carried out in the R statistical environment (R Core Team, 2012). In addition to the statistical approach, all samples were plotted on histograms using a Log_2 binning method whereby bin 1 equals the number of singleton species, bin 2 equals the number of species with 2–3 individuals, bin 3 equals 4–7, and so on (method 3 in Gray *et al.*, 2006a). Fits of the best distributions were analysed by eye in each case to check for any issues or peculiar fits.

We tested various mechanisms that may result in multimodal SADs. To test the effect of combining species subsets on the SAD, for each of the samples in which the PLN2 or PLN3 distributions provided the best fit, the octaves corresponding to the modes of the distribution were recorded and the species within these octaves labelled accordingly. For all samples in which the PLN2 distribution provided the best fit, a Welch's Two Sample t-test (0.05 significance level) was used to determine if the percentage of core, tourist, indigenous and high dispersal ability species significantly varied between the first and second mode. Furthermore, to determine whether the proportion of particular subsets within each mode were significantly different from a random sample of the same amount of species from the regional pool, a null distribution was created. For each of the modes within the samples in which the PLN2 provided the best fit the number of species representing the modal octave

class was recorded. This number of species was then randomly sampled without replacement from the regional pool and the proportion of core, tourist, indigenous, and high dispersal ability species calculated. The same method was used for mean body size. The number of iterations of this procedure was set at 10,000 for each mode. In order to make the regional pool sample more ecologically realistic, for any particular fragment the regional pool was restricted to data from the island in which the fragment was located. A 0.05 significance level was employed and thus an observed proportion falling in the top and bottom 2.5% of the sampled distribution was deemed to be significant. In addition to comparing the percentage of core and satellite species between modes within each sample, the PLN, PLN2 and logseries were calculated separately for the core subsets, and the PLN and logseries for the satellite subsets. It was then possible for each sample to determine if the core and satellite subsets were best modelled by separate distributions (sensu Magurran & Henderson, 2003; Gray *et al.,* 2005; Ulrich & Zalewski, 2006). The PLN2 distribution was not calculated for the satellite subsets as the reduction in sample size coupled with the large number of singletons made optimisation convergence problematic.

To determine if ENT provided an explanation for multimodality, we followed Vergnon *et al.,* (2012) and used body size as the basis of the niche axis along which assemblages were theorised to be organised. Body size frequency distributions were created for each of the samples using both untransformed and log(base ten) transformed body size values and any modes identified through visual inspection. Additionally, for each sample body size was plotted against species abundance (again, using untransformed and log transformed data) and all plots examined. Lastly, within samples that exhibited a significantly bimodal SAD, the mean body size of all species within the modal octaves was calculated and a Welch's Two Sample t-test used to assess whether the differences in body size between the two modes was significant across samples. Obviously, measures other than body size can be

used to construct a niche axis, and equally factors other than ENT can lead to multimodal body size distributions (see Holling, 1992), but we use body size as it has been recommended as an apt way of testing ENT (e.g. Segura *et al.*, 2011; Vergnon *et al.*, 2012)

Table 11.1 Site information for the different native forest fragments from the Azores used in the analyses, including the area and the number of arthropod individuals (N) and species (S) sampled within the four transects used to create the sample in each instance. Arthropods were sampled using a standardised pitfall trap and canopy beating methodology between 1999 and 2004. 'Island' relates to island within the Azores in which the fragment is located.

Fragment	Area (Ha)	Island	N	S
1. Cabeço do Fogo	36	Faial	5586	92
2. Caldeira do Faial	191	Faial	3356	80
3. Lagoas Funda e Rasa	240	Flores	5989	117
4. Morro Alto	1331	Flores	3107	89
5. Pico Pinheiro	74	São Jorge	7366	107
6. Topo	220	São Jorge	4939	102
7. Atalhada	10	São Miguel	3990	130
8. Pico da Vara	306	São Miguel	6204	118
9. Caldeira Guilherme Moniz	223	Terceira	6018	86
10. Pico Galhardo	38	Terceira	4608	111
11. Serra de Santa Bárbara	1347	Terceira	3131	89
12. Terra Brava	180	Terceira	4216	103

Table 11.2 Site information for the different Azorean islands used in the analyses. For each island the number of fragments that were combined to create the island dataset is provided along with the combined area, which is the summed area of the combined fragments. + This is not indicative of the actual number of native forest fragments that have been studied in the BALA project on a particular island, it is simply the number selected for this study. The number of arthropod species (S) and number of individuals (N) sampled in each island is also given.

Island	Number of Fragments+	Combined Area (Ha)	N	S
1 Faial	2	227	8942	125
2 Flores	2	1571	9096	141
3 São Jorge	2	294	12305	145
4 São Miguel	2	316	10194	178
5 Terceira	4	1788	17973	175

11.5 Results

In total, 56,338 individuals representing 399 unique species were analysed. Sample size varied from 3107 to 7366 individuals (mean = 4879) at the fragment scale (Table 11.1), and from 8942 to 17973 individuals (mean = 11702) at the island scale (Table 11.2).

11.5.1 Model comparison

The data showed significant evidence of multimodality in a number of samples. Six of the twelve fragments were best represented by the PLN2 distribution according to both BIC and AIC_c (Table 11.3). Three were best represented by the PLN3 distribution, again according to both model selection criteria. The remaining three fragments did not have a clear best model. For fragment 1 and 2 (both from Faial) the PLN2 distribution provided the best fit according to $AIC_{c.}$ However, according to BIC, the PLN2 and PLN, and PLN2 and logseries were both in the best model set (i.e. either the lowest BIC value or within a value of two of the lowest value) for fragment 1 and 2 respectively (Table 11.1). For fragment 3 (Flores island) the PLN provided the best fit according to BIC, but PLN, PLN2 and PLN3 were all within the best model set according to AIC_c (Table 11.1). At the island scale, two island samples were best fit by PLN2 (São Miguel, Terceira), and one by PLN3 (Flores), according to both criteria. The remaining two islands (Faial and São Jorge) had different best models if one used BIC or AIC_c (see Table 11.2). The LRT results are largely concurrent with model selection based on BIC and AIC_c (All LRT results are presented in Supplementary material Appendix 9, Table A9.1), indicating that the better fit of PLN2 and PLN3 compared to PLN is significant despite the increase in number of parameters. The difference in goodness of fit between PLN2 and PLN3 was often not significant (Table A9.1), suggesting that there is often no significant improvement in fit going from a bimodal to a trimodal model, although in all samples in which the PLN3 distribution provided the best fit according to the information criterion, the LRT results confirmed this.

For the samples in which PLN2 provided the best fit, one mode was apparent at the rare end of the distribution with the second modal octave generally being 16–32 or 33–64 individuals (see Fig. 11.2a for an example). Datasets best fit by PLN3 followed this pattern but with a third modal class at the more abundant end of the distribution (Fig. 11.2b). Nonetheless, visual inspection of the SAD histograms for certain samples indicated that the multimodal pattern was not always obvious despite the best model being either PLN2 or PLN3 (see Fig. 11.2c for an example). In sum, multimodality within the SAD, while not a universal pattern, does appear to be a common occurrence within natural Laurisilva forest arthropod communities in the Azores.

Table 11.3 Model selection results for arthropod SADs of twelve fragments of native Laurisilva forest in the Azores. Arthropods were sampled using a standardised pitfall trap and canopy beating transect methodology between 1999 and 2004. Each fragment consists of four transects. For each fragment the maximum log likelihood (MLL) is provided for each of the four SAD models: the standard Poisson lognormal (PLN), the two (PLN2) and three (PLN3) mode Poisson lognormal, and the logseries. The Bayesian Information Criterion (BIC) and Akaike's Information Criterion corrected for small sample size (AIC_c) calculated using the MLL are also presented. The PLN distribution has two parameters, PLN2 has five, PLN3 has eight, and the logseries one parameter. Fragment codes and abbreviated island name/initials relate to those in Table 11.1. Values in bold represent the best model for the corresponding fragment, for both BIC and AIC_c. Where several models are within an AIC_c or BIC value of two of one another, each is highlighted in bold.

Fragment	SAD Model	MLL	BIC	AIC_c
1 FAI	PLN	366.90	**747.36**	740.06
	PLN2	359.40	**745.92**	**731.78**
	PLN3	356.91	754.51	734.01
	Logseries	370.75	750.54	745.63
2 FAI	PLN	321.23	656.02	648.73
	PLN2	312.77	**651.84**	**638.70**
	PLN3	311.53	662.49	643.63
	Logseries	322.13	**653.01**	648.41
3 FLO	PLN	447.17	**908.65**	**900.54**
	PLN2	444.09	916.75	**900.94**
	PLN3	440.45	923.75	**900.57**
	Logseries	451.83	913.18	907.76
4 FLO	PLN	331.02	676.32	668.25
	PLN2	328.17	683.27	669.37
	PLN3	315.93	**671.19**	**652.48**

	Logseries	332.94	674.86	670.02
5 SJG	PLN	422.58	858.31	851.48
	PLN2	413.82	**855.67**	**840.48**
	PLN3	412.89	867.84	845.64
	Logseries	428.67	866.68	861.45
6 SJG	PLN	385.12	784.11	776.48
	PLN2		**777.03**	**762.16**
	PLN3	373.16	787.94	766.27
	Logseries	385.70	780.64	775.51
7 SMG	PLN	473.41	960.69	953.06
	PLN2	461.99	**953.18**	**936.66**
	PLN3	461.52	966.84	942.54
	Logseries	473.50	956.74	951.10
8 SMG	PLN	474.65	963.91	955.49
	PLN2	464.08	956.78	940.91
	PLN3	442.90	**928.74**	**905.47**
	Logseries	474.51	958.56	953.12
9 TER	PLN	364.18	741.82	734.63
	PLN2	356.03	738.78	725.12
	PLN3	307.86	**655.81**	**636.09**
	Logseries	363.55	736.02	731.25
10 TER	PLN	419.17	851.70	844.63
	PLN2	409.65	**847.56**	**832.11**
	PLN3	407.48	857.34	834.74
	Logseries	420.79	851.00	845.69
11 TER	PLN	211.20	436.52	428.62
	PLN2	197.20	**420.61**	**407.56**
	PLN3	194.54	426.23	410.55
	Logseries	208.35	424.96	420.91
12 TER	PLN	400.22	812.82	806.85
	PLN2	388.77	**803.76**	**790.71**
	PLN3	386.57	812.47	793.75
	Logseries	398.68	806.62	801.47

Table 11.4 Model selection results for arthropod SADs of five Azorean islands. Arthropods were sampled in the native Laurisilva forest using a standardised pitfall trap and canopy beating transect methodology between 1999 and 2004. Column and model abbreviations and the number of parameters of each model are as for Table 11.3. Values in bold represent the best model for the corresponding island, for both BIC and AIC_c. Where several models are within an AIC_c or BIC value of two of one another, each is highlighted in bold.

Island	SAD Model	MLL	BIC	AIC_c
1 Faial	PLN	400.22	814.34	806.68
	PLN2	388.77	**805.35**	**790.42**
	PLN3	387.57	818.67	794.69
	Logseries	398.68	**806.62**	801.47
2 Flores	PLN	568.17	1151.21	1142.52
	PLN2	560.22	1150.18	1133.07
	PLN3	550.77	**1146.15**	**1120.91**
	Logseries	572.87	1155.65	1149.83
3 São Jorge	PLN	582.15	1179.25	1170.47
	PLN2	573.19	**1176.28**	1158.99
	PLN3	567.85	1180.56	**1155.03**
	Logseries	593.13	1196.23	1190.35
4 São Miguel	PLN	766.36	1548.52	1538.84
	PLN2	756.69	**1544.99**	**1525.83**
	PLN3	754.65	1556.71	1528.28
	Logseries	775.09	1560.72	1554.25
5 Terceira	PLN	704.63	1424.66	1415.40
	PLN2	695.57	**1421.95**	**1403.65**
	PLN3	693.92	1434.05	1406.96
	Logseries	710.62	1431.51	1425.31

11.5.2. Post model selection

The percentage of core, tourist, and indigenous species was found to differ significantly between mode one (rarer species) and mode two (more common species) for samples best modelled by PLN2 or PLN3 according to a Welch's Two Sample t-test (Table 11.5); mode one at the rarer end of the distribution always had a higher percentage of satellite and tourist species and a lower proportion of indigenous species. These results were consistent at both the fragment and island scale (Table 11.5), although the difference in the proportion of

indigenous species across modes was not significant at the island scale. The percentage of species with high dispersal ability was also found to differ significantly between mode one and two, but the direction of the change was inconsistent among samples, that is, in some samples mode one had a higher percentage of species with high dispersal ability than mode 2, and *vice versa*. Thus, the effect of differences in dispersal ability as a mechanism for multimodal SADs in the Azores is unclear. No significant difference was found between any of the subsets between the second and third mode for samples best modelled by PLN3 (results not presented); although the degrees of freedom was reduced as PLN3 provided the best fit for fewer samples.

We found minimal evidence for the structuring of communities within each sample along a body size niche axis as outlined by ENT. Body size distributions were all distinctly unimodal, and the body size-abundance plots did not produce any apparent patterns. Mean body size was not found to significantly differ between mode one and mode two for datasets best modelled by PLN2 or PLN3 (Table 11.5).

While the randomly sampled null distribution varied between samples because they were drawn from different subsets of the archipelagic species pool and because of the number of species in each modal octave class, a rough rule of thumb in relation to the proportion of core species was that the 2.5^{th} and 97.5^{th} percentiles were located at 5% and 35% core species. Thus, at both the fragment and island scale it was universally the case that there were more core species in the more abundant mode than expected by chance. For the rarer mode, there was never a lower proportion of core species than expected by chance at the fragment scale, but at the island scale three of the four samples in which PLN2 provided the best fit had a significantly lower proportion of core species than expected by chance. At the fragment scale, the rarer mode always had a higher proportion of tourist species than expected by chance, although this difference was significant ($P < 0.05$) in only five of the eight cases. No

significant differences in proportion of tourists were found at the island scale. In relation to the more abundant mode, two fragments and three islands had a lower proportion of tourist species than expected

by chance. Considering the proportion of indigenous species, for two fragment samples and one island sample there were more indigenous species in the more common mode than expected by chance, and in no case was there significantly fewer indigenous species in the rarer mode. No consistently significant results in terms of the proportion of species with high dispersal ability or the mean body size were found.

Table 11.5 Welch's Two Sample t-test results for the composition of the different modes within the SAD, with the degrees of freedom (df) and the corresponding *P* value, for both the fragment and island samples, and the direction of change. The values only relate to samples in which the PLN2 distribution provided the best fit, that is, there were two distinctive modes. The direction of change relates to whether the proportion of a subset was consistently higher or lower in the more common mode across samples. For instance, considering dispersal ability, in some samples the rarer mode had a higher proportion of high dispersal ability species and in other samples the opposite pattern was observed. Values in bold represent results significant at the 0.05 level.

Fragments				
Subset	Direction of change	t	df	*P*
Percentage of tourist species	Lower in more common mode	7.76	20	**0.000**
Percentage of core species	Higher in more common mode	-13.43	19	**0.000**
Percentage of high dispersal ability species	Mixed	2.82	20	**0.011**
Body Size	NA	-0.38	15	0.708
Indigenous species	Higher in more common mode	-1.07	17	**0.002**
Islands				
Subset	Direction of change	t	df	*P*
Percentage of tourist species	Lower in more common mode	4.59	8	**0.002**
Percentage of core species	Higher in more common mode	-5.26	4	**0.005**
Percentage of high dispersal ability species	Mixed	2.32	8	**0.049**
Body Size	NA	-1.38	8	0.204
Indigenous species	Higher in more common mode	-2.36	6	0.059

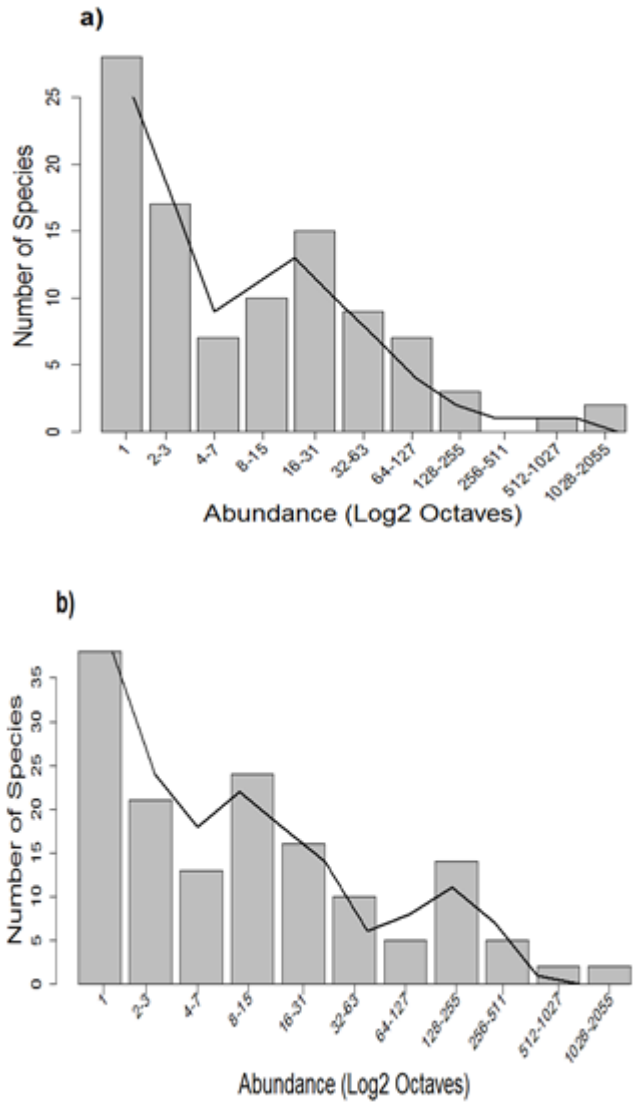


Figure 11.2 Exemplar multimodal species abundance distributions with fit lines for Azorean arthropods sampled using a standardised pitfall trap and canopy beating sampling methodology. Data were gathered from 1999–2004. a) Fit of the PLN2 distribution (black line) to the data from fragment 1 (located on the island of Faial). b) Fit of the PLN3 distribution (black line) to the data from island 2 (Flores). c) Example of a SAD for which the PLN2 provides the best fit according to both information criterion (BIC and AIC_c) but for which the bimodal pattern is not immediately striking. Data are from island 4 (São Miguel). Island samples were created by pooling the fragment data on each island. In each plot abundances are grouped into Log_2 octaves: octave 1 contains the number of species with 1 individual, octave 2 the number of species with 2-3 individuals, octave 3 the number of species with 4-7 individuals, and so forth.

Re-running the model selection using the core and satellite subsets separately revealed mixed results (BIC and AIC_c values for all models and for each sample are presented in Supplementary material Appendix 9, Table A9.2). For the core subset at the fragment scale it was often unclear which of the logseries or PLN provided the best fit as in only seven samples did a particular distribution provide the best fit according to both BIC and AIC_c using a minimum difference in each criterion of two. Nevertheless, apart from the single fragment where the PLN2 distribution was the best model, and a further two fragments where the PLN2 distribution was within the best model set according to AIC_c, the good fit of the PLN2 more generally, and thus the presence of multiple modes, was not apparent when

focusing purely on the core subset of species. For the satellite subset the PLN performed best according to both criteria for all but one fragment. Results at the island scale were similar. Considering the core subset, the PLN provided the best for all island samples according to BIC and in four out of five instances according to AIC_c. The PLN performed best in all five instances for the satellite subset, according to both criteria.

That the PLN provided a better fit than the logseries for the satellite species in most samples is intriguing as analysis of the SAD frequency distributions reveals a logseries type pattern in the majority of instances. Fig. 11.3 illustrates the pattern exhibited in most samples, although the exact from of the relationship varies between samples. The complete assemblage (Fig. 11.3 a) exhibits two distinct modes, one of very rare species represented by only one individual in the sample, and one of relatively more common species. Deconstructing into core and satellite subsets indicates that the core subset exhibits a more lognormal type pattern, with the mode corresponding to the common species mode in the complete assemblage SAD. Accordingly, the satellite subset SAD exhibits a more logseries pattern (but is best fit by the PLN; Fig. 11.3 c). It is clear then, for this sample at least, that combining the core and satellite subsets within the complete assemblage results in two distinct modes in the SAD. Results were largely consistent between the two sets of core/satellite classification methods and we are thus confident that our choice of 50% as the grouping threshold did not affect the results. For example, Fig. 11.3 (d-f) shows the plots for the same data as Fig. 11.3 (a-c) but using the three group classification. It is evident that it is the "25% or fewer group" which is generating the first mode in the SAD of the whole assemblage (Fig. 11.3a), and that the 76-100% group is generating the second mode.

11.6 Discussion

The performance of Poisson lognormal SAD models that incorporate multiple modes were compared with the traditional one mode PLN and logseries distributions for twelve forest

fragments, and the combination of fragments within five islands. Multiple modes were found in a large number of samples, with the PLN2 distribution most prevalent, followed by the PLN3 distribution. Consistent with what we have termed the amalgamation hypothesis (see Introduction) it appears that the amalgamation of core and satellite, and indigenous and introduced subsets leads to the observed multiple modes. The presence of tourist species, which in many (but not all) instances are also satellite species, and which are present in larger abundances in more anthropogenic land uses, inflates the number of rare species in the samples and thus leads to the emergence of a rare mode in the SAD.

11.6.1 A deconstruction approach

Combing assemblage deconstruction within a multimodal SAD framework has been put forward as a promising way of advancing SAD research (McGill *et al.*, 2007). Nonetheless, our finding that grouping together core and satellite species into one assemblage results in two distinct components of the SAD is not unique. For instance, some studies have reported different SAD patterns for core and satellite subsets using time series data (e.g. Magurran & Henderson, 2003; Ulrich & Ollik 2004). Ulrich & Zalewski (2006) have also found different SAD patterns for core and satellite groups, in their case using a very similar methodology to the present study, determining group membership based on the proportion of samples across space in which each species was recorded (and see Unterseher *et al.*, 2011 for an example using molecular taxonomic units).

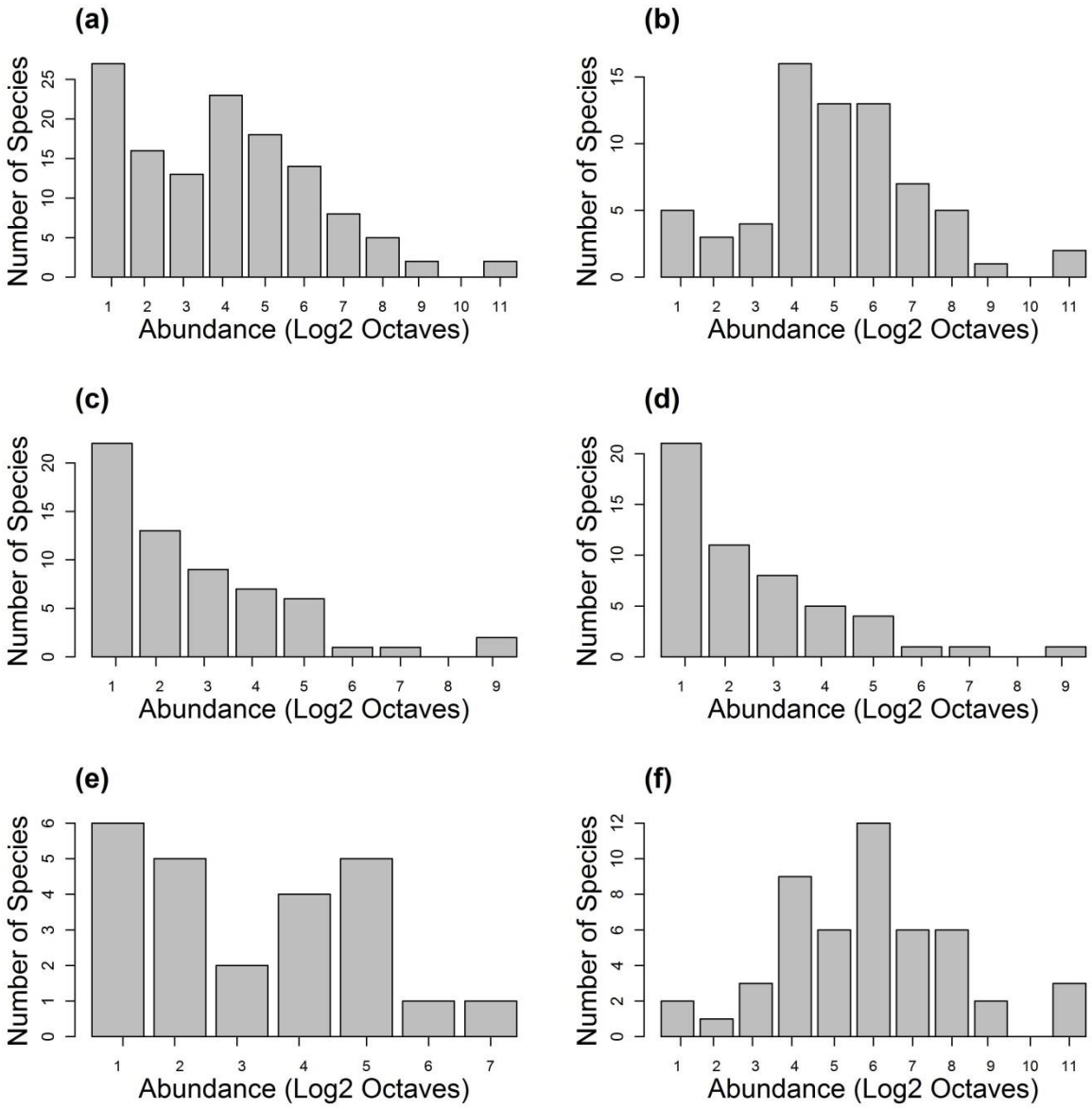


Figure 11.3 The SAD of arthropods from the island of Faial in the Azores (as Table 11.2, and Fig. 11.1). Species were sampled using a standardised pitfall trap and canopy beating sampling methodology from 1999–2004. Plots correspond to the SAD of: a) the whole assemblage, deconstructed into, b) core species and, c) satellite species. Two modes are clearly present in (a) which correspond well with modal octaves in (b) and (c). d-f correspond to the three group core and satellite approach, whereby d) represents species found in 25% or fewer fragments, e) 26-75% of fragments and f) 76-100% of fragments. In each plot abundances are grouped into Log_2 octaves: octave 1 contains the number of species with 1 individual, octave 2 the number of species with 2-3 individuals, and so forth

.

Our findings differ from the aforementioned studies in that the satellite subsets were best modelled by a PLN distribution and the core species by a mixture of PLN, PLN2 and logseries distributions. Thus, in certain instances both the core and satellite subsets were best modelled by PLN distributions but with different modal octaves, a pattern similar to that

observed by Ugland & Gray (1982) in marine benthic communities. Regardless of which distribution best fits the core subsets, for the bimodal samples it is clear that the proportion of core species is significantly higher in the more common mode. Examining the concept of core species reveals that this finding is to be expected. Core species, as classified herein, are those which are theorised to be the constituent members of the community, organised according to traditionally invoked deterministic niche-based processes, such as competition and predation (Begon *et al.*, 1990; Ulrich & Zalewski, 2006). As such, it intuitively makes sense that they will, by and large, obtain higher abundances than the satellite species, which are predominantly immigrants from outside the community (discussed below). It is also possible that our method of classifying species as core and satellite has influenced our findings. While our method of classification is independent of species abundance at the fragment scale in that a species classified as core or satellite can theoretically have any abundance in a particular fragment, and we found that our 50% threshold was robust, a positive occupancy–abundance relationship is known to exist for a variety of taxa and systems, including for Azorean arthropods (Rigal *et al.*, 2013). As we have classified species based on incidence, it is possible that the core species have a higher abundance simply as a result of this relationship. However, the large overlap in identity between satellite and tourist species indicates that many satellite species obtain much larger abundances in other land uses and we are confident that our results are not simply a result of a positive occupancy–abundance relationship. Equally, as we have pooled data from individual fragments to create our island datasets, there is a chance that a degree of circularity exists in our island datasets, i.e. whereby the classification of a species as core is not independent of its abundance. For example, if an island contained two thirds of the fragments in our dataset, the pooled abundance of any core species would have to be greater than one, and thus the species could not be represented in the singleton octave of the SAD. In reality, this is only an issue if any

islands contain a large proportion of the fragments used in the core-satellite classification. As the island in our dataset which contains the most sampled fragments (Terceira; Table *), only contains four fragments, it is unlikely that this circularity is driving the observed patterns at the island scale.

While we found no evidence of communities being structured along a body size niche axis, this does not disprove ENT in respect to our study system. We only focused on body size as a measure of the niche, which while recommended by Vergnon *et al.,* (2012), is not the sole niche axis along which ecological communities could be structured (Scheffer & van Nes, 2006; Vergnon *et al.,* 2012). As no additional measures of a niche axis were tested, all we can say with surety is that ENT based on body size as an indicator of the niche is not the cause of the observed multimodal SADs in our data. It is also possible that focusing on the body size of arthropods in general obscures any general pattern as this grouping comprises a mix of various taxa, e.g. spiders, beetles, and Lepidoptera, each with a different body size plan or 'bau plan'.

It has been postulated that satellite species dynamics are mainly controlled by stochastic processes, specifically random dispersal, as opposed to the niche-based mechanisms hypothesised to be structuring core species (Magurran & Henderson, 2003; Ulrich & Ollik, 2004). Our results seemingly back up this line of argument. In the Azorean invertebrate arthropod community it appears that species classified as satellite species are predominantly, but not universally, also tourist species. That is, they are not simply species which are only occasionally found in a set of samples (i.e. satellite species), but are also species which are more abundant in other non-native habitats (i.e. tourist species; Borges *et al.,* 2008). Thus, it is the immigration of small numbers of individuals from these different habitats into the native forest that results in the mode of rare species. This is a key finding. While other studies have focused on the core/satellite dichotomy and assumed that satellite

species have different habitat requirements, we have gone a step further and shown that the presence of tourist species in an assemblage significantly affects the signal in the SAD.

Our data derive from native forest fragments on the Azores, a system where there has been large scale conversion of native forest to semi-natural and intensive pasture alongside a long and continuing process of anthropogenic introduction of non-native species (Borges *et al.,* 2008; Cardoso *et al.,* 2010). Exotic forestry plantations are also a common feature of the landscape and, currently, native forest represents only a small amount of the total Azorean land use (Cardoso *et al.,* 2009). These changes, in combination, mean that many arthropod species now found in the Azores are better adapted to the more anthropogenic habitats. It is thus evident that the concepts of satellite species and tourist species are interlinked. The species sampled in only a few fragments, i.e. the satellite species, are largely immigrants from land uses outside the fragment and will thus naturally be represented in our samples by only a few individuals (see also Gray *et al.*, 2005, 2006b). However, linked to the issues described above, it is possible that our classification of tourist species influenced our results. We classified a species as tourist if it possessed a greater abundance in any of the three anthropogenic habitats relative to the native forest. Thus, implicit in this classification is our assumption that species should be abundant in one habitat (e.g. native forest) and rare in the others (e.g. the more anthropogenic habitats). However, this is not necessarily the case and it is possible that a positive relationship between abundance in the different habitat types exists, such that certain species are abundant in both sets of habitats and *vice versa*. Obviously, this positive relationship will characterise certain species. For instance, in our data roughly 10–20% of tourist species are core species, depending on the sample in question, and these are likely to be predominantly habitat generalists; species found in large abundances in multiple habitats (Shmida & Wilson, 1985; Borges *et al.,* 2008; Cardoso *et al.,* 2009, 2013). Nonetheless, in general we do not think this assumption is problematic in our analysis, as

overall we found a negative correlation between abundance in the native forest and max abundance in the additional habitats (both variables log transformed; Pearson's product-moment correlation: r= -0.23, P=0.008).

It is also the case that there is a higher proportion of introduced species in the rarer mode of the bimodal SADs. The Azorean invertebrate arthropod fauna includes a vast number of introduced species (Borges *et al.,* 2010), evidenced by the observation that even in the common mode of the SAD up to 40% of the species in the modal octave were classified as introduced. For instance, one of the most widespread introduced species in man-made habitats, the millipede *Cylindroiulus propinquus,* was also successful in invading native forest, where it is now abundant. It has been shown that introduced species are integrated alongside indigenous species in the contemporary Azorean arthropod communities such that they are mostly undetectable through the study of abundance–occupancy plots (Rigal *et al.,* 2013). This implies that current island faunas with a high proportion of introduced species may have an inflation of rare species in native habitats due to source–sink dynamics (see also Borges *et al.,* 2008).

11.6.2 Distinguishing between two and three modes

While the grouping of core and satellite, and indigenous and introduced species in a sample appears to explain bimodal SADs, it does not provide an answer as to why certain samples were best modelled by the PLN3 distribution. For instance, taking fragment 9 (see Table 11.3) as an example it can be seen that PLN3 provides the best fit for the whole assemblage (Table 11.3), and PLN2 the best for the core species (Supplementary material Appendix 9, Table A9.2a). Inspection of the frequency distributions reveals the modal octaves for the core species are 8–15 and 64–127 species, which is largely congruent with the pattern for the whole assemblage (although without the mode of very rare species), while the satellite species have a large number of singletons. Hence, the two more common modes of the whole

assemblage are both mainly comprised of core species, with few to zero tourist species, meaning that there must be a further causal mechanism driving this pattern. The grouping of species into ecological guilds may add some explanatory power (see Borges *et al.,* 2008) and it is also possible that in such communities there are three distinct groups of species, as postulated by Ugland & Gray (1982; see also Borges *et al.,* 2008). Further research is needed to elicit the structuring mechanism(s) governing these multimodal groupings.

We found that for certain samples distinct multiple modes were not particularly evident visually (e.g. cf. Fig. 11.2c), although they were detected to be multimodal in our analyses. Thus, it may be the case that many well studied communities are in fact multimodal: it is simply that appropriate statistical methods have not been applied to detect this (see also McGill *et al.,* 2007; Vergnon *et al.,* 2012). As the R code is now available (Dornelas & Connolly, 2008) to test multimodal PLN distributions we call on researchers who are exploring SAD patterns in different datasets to statistically test for these patterns, even if they are not immediately apparent visually.

In sum, the amalgamation of samples through time and space can lead to a bimodal – or even trimodal – SAD within which there is a discernible group of occasional species each represented by only a few individuals, leading to a distinct mode at the rare end of the SAD. In the Azorean Laurisilva forest arthropod community many of these satellite taxa are introduced species best adapted to non-native land uses but occasionally present and sampled in the native forest. Thus, deconstructing the full assemblage into various subsets (Marquet *et al.,* 2004; Borges *et al.,* 2008) can aid in elucidating the underlying patterns and we therefore recommend this approach for future analyses of SADs. Nonetheless, it is imperative that care is taken in determining ecologically rational deconstructions as simply classifying all relatively rare species from a sample as satellite species, for example, will obviously result in a change in the form of the core and satellite SADs. Multimodal SADs may be more common

than originally perceived and we advocate testing for them using appropriate rigorous statistical methods.

CHAPTER 12

Examining variations in incidence functions across the range

12.1 Preface

This chapter has not been written up as a paper and is thus only loosely structured as such. A comprehensive summary of incidence functions, and introduction to the analyses in this chapter are given in Chapter 2 (section 2.7). Due to word constraints, this introduction and literature review will not be repeated here. Rather, I will provide a brief overview of the study alongside the study hypotheses. In addition, this chapter includes a reduced discussion section as the results and limitations of the study design are discussed in Chapter 13.

As discussed in Chapter 2, I use the term incidence function solely to refer to models constructed using island area as the independent variable.

12.2 Introduction

The utility of incidence function models in conservation applications rests on the assumption that a given species' incidence function curve does not vary to any substantial extent across space. The few studies that have attempted to answer this question have generally found some degree of spatial variation in incidence function model parameters (e.g. Watson *et al.*, 2005). This chapter seeks to analyse variations in avifaunal responses to forest fragmentation, specifically through ascertaining how various bird species respond to habitat fragmentation at different points within their range, in terms of area sensitivity. Specifically, I evaluate whether a species' incidence function curve is constant at different point throughout the species' range. Four fragmented European landscapes representing different points within a target group of species' overlapping ranges, and containing multiple forest fragments along a patch-area gradient, have been surveyed for birds over a two year period. For each target species, incidence function curves will be compared by comparing the critical island area value (CIV; i.e. the island area value which corresponds to a 50% probability of occurrence,

according to the model; see Chapter 2) across landscapes. Based on the core-abundance hypothesis (see Chapter 2), it is hypothesized that species will respond to fragmentation in contrasting ways at the range core as opposed to the range margin. The null and alternative hypotheses are:

H0 – For a given species, there is no systematic variation in the critical island area between landscapes located in the range core and the range edge.

H1 **–** For a given species, the critical island area will be lower (i.e. a smaller area) for landscapes located in the range core than those in located the range edge.

12.3 Methods

12.3.1 Data collection

The full sampling methodology and summary of the original target species is provided in Chapter 3. In brief, six target bird species were sampled in sets of forest fragments (n = approximately 40 in each case) of varying area in four landscapes: two range core (France and south UK), and two range edge (south Spain and Norway) landscapes. Birds were sampled using a point count methodology (see Chapter 3), and each fragment was sampled two-three times. As every fragment was not sampled three times, only the first two repeats were used in these analyses. For each fragment the two repeats were merged to create a single presence-absence value. Thus, for each target species the final dataset comprised four vectors of presence-absence data relating to each of the four landscapes.

During preliminary data analysis a problem arose due to the fact that binary generalised linear models (GLMs) using a logit link function (outlined below) are not applicable when the proportion of presences is much higher than the proportion of absences, or *vice versa* (Zuur *et al.*, 2009). Whilst it is possible to use different link functions with such data (e.g. the clog-log link), the issue then arises that the incidence function curve takes a different shape (i.e. Zuur *et al.*, 2009) and it is difficult to compare curves and CIVs. It is also

the case that binomial GLMs do not give accurate parameter estimates when fitted to data with very few presences. As this was much more of an issue with the range edge landscape data, ignoring it could have resulted in biased results. Thus, for statistical reasons it was necessary to reduce the number of target species from six (see Table 3.1) to four. The condensed set of species comprised: wood pigeon (*Columba palumbus*), blackbird (*Turdus merula*), robin (*Erithacus rubecula*), and blue tit (*Parus caeruleus*).

12.3.2 Statistical analysis

As outlined above, my incidence function models were purely focused on the effect of island area. This approach is only appropriate when area is an important or significant predictor variable. Thus, before fitting the area incidence function models I fitted a global binomial GLM, with the presence-absence of a target species as the response and the variables described in Chapter 3 (e.g. habitat complexity) as the predictors. Colinearity between the predictors was tested using variance inflation factors. A number of the predictors were correlated with each other, and thus a reduced set was used for these analyses: area, isolation, vertical vegetation density and habitat complexity. For each target species, the global model was fitted to the data from each of the four landscapes separately. A simple stepwise procedure was employed using the step function in R to select the best model according to AIC. Whilst more rigorous model selection procedures could have been used (e.g. those in Chapters 5-6), this model fitting was simply done as a check to determine whether area was an important variable explaining the presence of the target species in the sampled fragments (i.e. the model was not used for prediction etc.). Any target species for which area was not a significant variable in the global model for any of the four landscapes was flagged, as interpreting an area incidence function model in such instances does not make sense (Watson *et al.*, 2005).

Incidence function models were fitted as a GLM from the binomial family and with a logistic link function, whereby the presence-absence of a species in a set of patches (from a single landscape) was used as the response variable, and the area of each island was used as the predictor variable (e.g. Helzer & Jelinski, 1999; Oksanen, 2004; Watson *et al.*, 2005). Whilst the response variable represents a binomial outcome (i.e. a species is either present or absent), the binomial GLM returns a probability, which can be interpreted as the probability that a species is present on an island of a given size. For each target species, I fitted the GLM four times: once to the data from each of the four landscapes. For each model fit I recorded the effect of area on the response, McFadden's pseudo-R^2 (it is not possible to compute the standard R^2 statistic for binomial GLMs) and the critical island area value (CIV). The CIV was taken as the island area value corresponding to a 50% probability of occurrence, according to the model (see Fig. 2.5; cf. Watson *et al.*, 2005; Hanski, 2010). Occasionally, the lowest fitted probability of occurrence was greater than 0.5. That is, for the smallest island in that landscape, the target species had a greater probability of occurrence than 0.5. Thus, calculating the 50% CIV in such instances resulted in values which did not make ecological sense (e.g. -1). In these instances, the CIV was taken as <0.01ha. In sum, for each of the target species four CIVs were returned: two core landscape and two edge landscape values. To my knowledge, it is not possible to compare these values in a statistical sense as sample size is so low (i.e. two values in each treatment). Thus, as a first pass the core CIVs were simply compared with edge CIVs to see which were higher or lower. Second, to enable a more robust determination of the effect of range location on the incidence function, another binomial GLM was fitted to each target species' presence-absence data. However, instead of fitting a model to each landscape separately, the data from the four landscapes were combined into one data frame. An additional binomial categorical variable was created which stated whether a fragment was from a core or an edge landscape (location). Thus, the new

data frame had one response (presence-absence in each of the fragments), and two predictors (fragment area and the binomial range location classification). The GLM was fitted to these data, including a term for the interaction between area and location. If the interaction term was found to be significant, I concluded that range location had an effect on the area incidence function. More formal model selection procedures were not required due to the very small number of predictors (i.e. two).

To enable a visual comparison of the four incidence functions of each target species, I plotted each using the predicted values from the four model fits and the ggplot2 R package (Wickham, 2009).

12.4 Results

Overall, the incidence function analysis results were inconclusive. Considering the global GLMs, area was generally one of the most important predictor variables out of those considered (Table 12.1). With one exception, area was always included in the best model (Table 12.1). That being said, the effect of area in the global model, whilst always positive, was not always significant (Table 12.1).

The CIV values were lower in the two core landscapes than in the two range edge landscapes for wood pigeon and robin (Table 12.2). For blackbird and blue tit the two lowest CIV values corresponded to a core and an edge landscape (Table 12.2). The CIV of blue tit in the French landscape (range core) was the highest observed for any target species. Examination of the incidence function plots revealed similar findings to the CIV analysis (see Fig. 12.1).

Table 12.1 The results of the global GLM analysis. For each of the four target species, the presence-absence of the species in fragments in one of the four landscapes was used as the response variable, and island area (Ar.), isolation (Is.), habitat complexity (HC) and vertical vegetation density (VD) were used as the predictors. The effect sizes and *P* values given in the table relate to the fit of the global model (i.e. all predictors). A stepwise procedure was then used to select the best model according to AIC. If island area was a predictor in the best model, a 'Yes' value was placed in the 'Best' column. McFadden's pseudo-R^2 value is also given for each global model fit.

Wood pigeon										
	Ar. (Effect)	Ar. (*P*)	Is.(Effect)	Is.(P)	HC (Effect)	HC (P)	VD (Effect)	VD (P)	R2	Best
France	0.58	0.02	<-0.01	0.90	-0.18	0.90	0.04	0.48	0.40	Yes
Spain	0.08	0.07	<-0.01	0.47	-0.09	0.47	0.01	0.65	0.16	Yes
UK	0.20	0.09	<-0.01	0.98	-0.07	0.98	0.05	0.80	0.36	Yes
Norway	0.14	0.02	<-0.01	0.56	0.31	0.56	0.05	0.31	0.35	Yes
Robin										
	Ar. (Effect)	Ar. (*P*)	Is.(Effect)	Is.(P)	HC (Effect)	HC (P)	VD (Effect)	VD (P)	R2	Best
France	0.28	0.05	<-0.01	0.73	-0.49	0.73	0.03	0.10	0.27	Yes
Spain	0.25	<0.01	-0.01	0.18	0.37	0.18	0.06	0.21	0.41	Yes
UK	0.22	0.06	-0.01	0.48	0.43	0.48	0.04	0.15	0.44	Yes
Norway	0.05	0.04	-0.01	0.34	0.20	0.34	0.02	0.50	0.28	Yes
Blackbird										
	Ar. (Effect)	Ar. (*P*)	Is.(Effect)	Is.(P)	HC (Effect)	HC (P)	VD (Effect)	VD (P)	R2	Best
France	0.37	0.24	<-0.01	0.47	0.17	0.47	0.08	0.60	0.29	Yes
Spain	0.09	0.26	<-0.01	0.58	0.42	0.58	-0.03	0.10	0.18	Yes
UK	0.38	0.16	-0.04	0.07	1.09	0.07	-0.01	0.14	0.68	Yes
Norway	0.12	0.01	<-0.01	0.90	0.44	0.90	0.02	0.18	0.36	Yes
Blue tit										
	Ar. (Effect)	Ar. (*P*)	Is.(Effect)	Is.(P)	HC (Effect)	HC (P)	VD (Effect)	VD (P)	R2	Best
France	0.01	0.55	<-0.01	0.74	0.19	0.74	-0.06	0.39	0.12	No
Spain	0.19	0.17	<-0.01	0.19	-0.05	0.19	-0.04	0.83	0.26	Yes
UK	0.41	0.24	0.01	0.54	-0.11	0.54	0.03	0.73	0.28	Yes
Norway	0.06	0.04	<-0.01	0.76	0.28	0.76	0.03	0.32	0.25	Yes

Table 12.2 The critical island area values (in ha) for four target bird species, across four landscapes.

Landscape	Wood pigeon	Robin	Blackbird	Blue tit
France	2.17	<0.01	<0.01	38.17
Spain	14.53	9.93	<0.01	<0.01
UK	1.11	3.42	2.19	<0.01
Norway	7.49	28.5	12.43	26.44

When the presence-absence data from each of the four landscapes was combined for each target species, the GLM results were again inconclusive. The effect of area was significant for all species except blue tit, but the effect of the location variable (i.e. a categorical variable outlining whether a fragment was from a range core or range edge landscape) was only significant for robin (Table 12.3). The term representing the interaction between area and location was always negative, but the interaction was never significant (Table 12.3).

Table 12.3 The combined binomial GLM results. For each of the four target species, the data from the four landscapes was combined and a new variable which classified each data point as either belonging to a range core or a range edge landscape. The three predictor terms used in the GLM were island area, landscape location (i.e. core or edge) and the interaction between area and location. The effect and *P* value for each term is given along with McFadden's pseudo-R^2 value. Significant *P* values are highlighted in bold.

	WP	ROB	BB	BT
Area (effect)	0.27	0.18	0.26	0.03
Area (*P*)	**<0.01**	**0.01**	**0.02**	0.14
Location (effect)	-0.57	-1.14	-0.33	-0.26
Location (*P*)	0.34	**0.04**	0.59	0.54
Interaction effect	-0.17	-0.11	-0.18	0.01
Interaction (*P*)	0.08	0.11	0.12	0.71
R^2	0.28	0.27	0.20	0.06

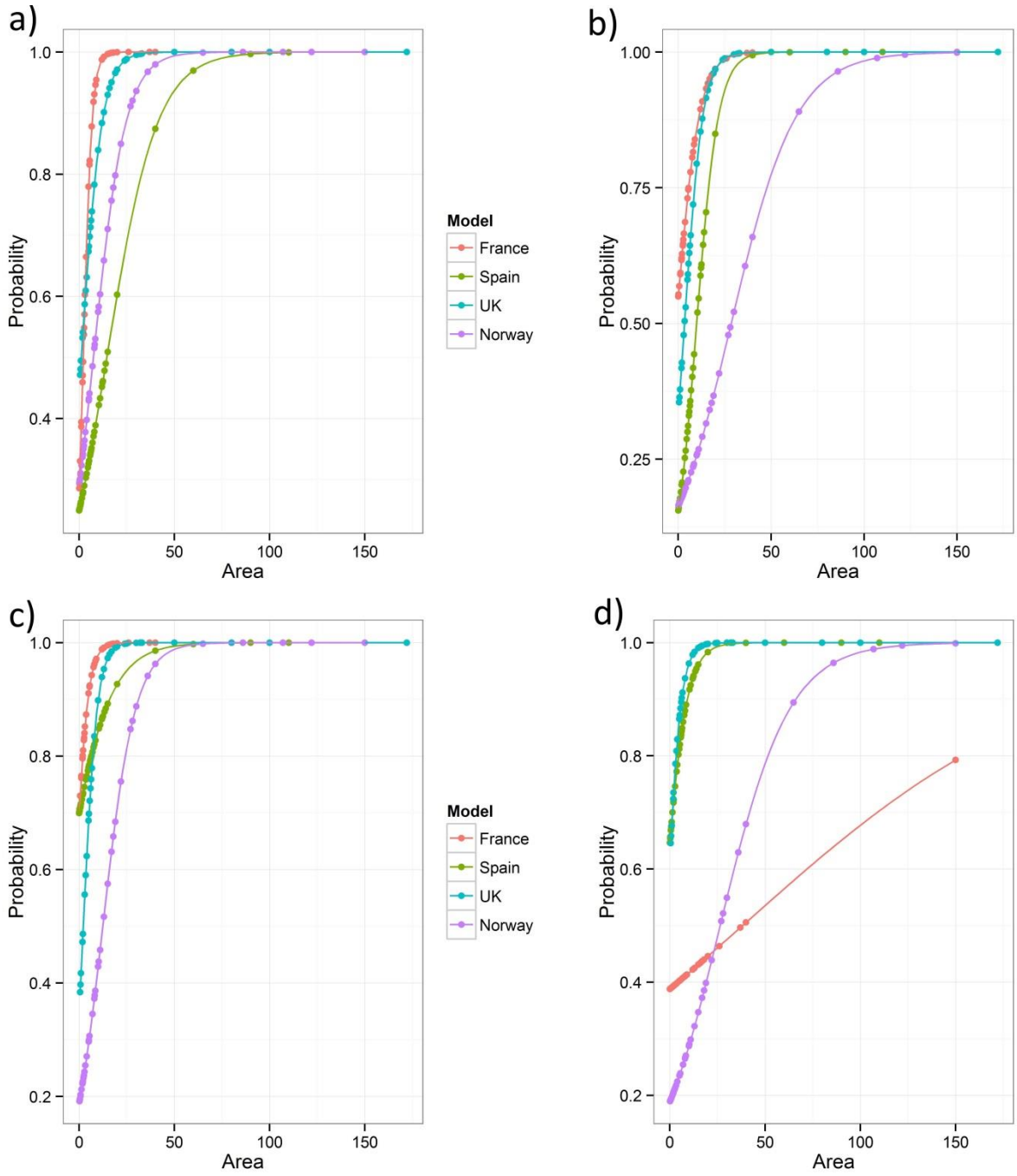


Figure 12.1 The incidence function plots of four target species: a) wood pigeon, b) robin, c) black bird, and d) blue tit. In each plot, the incidence function for each of the four landscapes is plotted (the legend is provided in the figure). Note that the y-axis does not start at zero.

12.5 Discussion

Overall, my incidence function results were inconclusive. For two species (wood pigeon and robin) the CIV occurs at a lower island area in the range core landscapes compared to the range edge landscapes (Table 12.2). Thus, based solely on these analyses (i.e. the CIV values

and plots) one could tentatively reject the null hypothesis (H0) and accept the alternative hypothesis (H1). However, whilst the effect of area in the combined GLMs was significant for both these species, the effect of the location variable was only significant in the case of robin, and the interaction between area and location was not significant for either species. The effect of the location variable and interaction term was negative for three of the four target species (Table 12.3), and so one may infer that moving from the range core to the edge results in a decreased probability of presence for a given fragment area. However, the lack of significance of these terms and the positive effect of the interaction term for blue tit again makes the results difficult to interpret. If the effect of island area on the incidence of the target species differs at the range edge compared to the range core we would expect the interaction term in the combined GLMs to be significant. Examination of the plots highlights why the location and interaction terms were often non-significant. Whilst the two range edge landscapes do have lower CIV values than the two range core landscapes for both wood pigeon and robin (Fig. 12.a, b), it is clear that with the exception of the Norwegian landscape in the case of robin (Fig. 12b) the range edge curves are not that different from the range core curves.

The CIV values for the two remaining target species (blackbird and blue tit) were not lower in both the core landscapes relative to the range edge landscapes. Furthermore, the effect of area was not significant for blue tit in the combined GLM analysis (Table 12.3). Thus, for these species it is not possible to reject the null hypothesis (H0) that the CIV values vary systematically across the range.

The GLM analyses are a more robust approach than simply comparing CIV values and incidence function plots. Analysis of incidence function plots can be misleading as plotting a line through the model's predicted values almost always generates a smooth curve, even if the effect of area is not significant or explains little variance. For example, for my

target species the combined models generally explained a very small amount of variation in the presence-absence of each target species (Table 12.3): the pseudo R^2 value was never above 0.3, and for blue tit it was only 0.06! As such, caution should be employed when using CIV values in future studies. Whilst a number of studies have compared CIV values, or some variant, few studies report the pseudo R^2 value of the incidence function models (Watson *et al.,* 2005). A model that explains less than 20% of the variance in the response variable of interest should not be really be used as the basis of any reasonable inference (see Burnham & Anderson, 2002). van Dorp & Opdam (1987) also found that for certain bird species in the Netherlands, there was no significant relationship between fragment area and presence, and these authors also caution the use of incidence function models in such cases.

The peculiar shape of the incidence function for blue tit in the French landscape (Fig. 12d), and possibly also the low pseudo R^2 value in the combined GLM analysis for the same species, is ostensibly due to the fact that the species was not sampled in the largest fragment during the first two rounds of sampling. This could be because the species was truly absent, or it could be simply that the species was missed. The largest fragment in the French landscape was roughly 150ha and even with a large number of point counts, it can reasonably be assumed that it would have been easy to miss a species during sampling. The issue of species detectability is discussed below, but this finding highlights a potential issue with the incidence function method. Namely, that it appears to be sensitive to (potential) false absences in particular influential data points/islands (see Mackenzie *et al.*, 2006).

The lack of conclusive findings makes the identification of underlying processes relatively meaningless. In Chapter 2, I argued that there were two primary explanations that could result in a larger CIV value in range edge landscapes relative to core landscapes. Both explanations were based on the core-abundance hypothesis: the idea that species have higher incidence and are more abundant in their range core, and that abundance and incidence

decreases towards the range margins. First, for a given species niche conditions are predicted to be more favourable in the range core and decrease towards the range margins. Thus, according to this hypothesis, extinction rates are elevated at the range margins and a species has a lower probability of presence in any given sized fragment (Brown, 1995; Sagarin & Gaines, 2002; Gaston, 2003). Second, as abundance is predicted to decrease towards the range margins, species may be absent from smaller patches simply due to chance, and not as the result of any niche-driven extinction processes. According to random placement models, species are distributed across a set of fragments in a landscape as a function of the species total abundance and the area of each fragment; such that individuals are more likely to be present in larger than smaller fragments (Coleman, 1981; van Dorp & Opdam, 1987; Wright, 1991; Gaston, 2003). Thus, CIV values would again be larger in range margin landscapes. For wood pigeon and robin it does appear, at least to a small degree, that incidence probability is reduced in small fragments at the range margin relative to the core (Fig. 12.1). However, the lack of explanatory power of the incidence function models, coupled with the non-significant effect of the location variable and interaction term means that neither of the above hypotheses can be unequivocally proven for either species. In fact, a more parsimonious conclusion may be that the hypothesised range effect and core-abundance hypothesis are not applicable to any of my four target species (see also Blackburn *et al.*, 1999; Sagarin & Gaines, 2002).

The limitations regarding the sampling methodology and study design are discussed in more depth in a standalone section (13.4.2). However, it is worth briefly summarising some of these issues here. First, I was only able to collect suitable data for a small number of species. Logistical constraints meant Europe was the only region in which I was able to sample, and the relatively low bird diversity in Europe meant there were few species with overlapping ranges which were suitable for inclusion in this analysis. It is also the case that

most of these target species are not forest specialists, and thus generalising my results to more specialist species should only be undertaken with a degree of caution. Whilst the inclusion of more forest specialist species would have been preferable, it was not possible to select a set of overlapping ranges in Europe that included many specialists. Furthermore, it is likely that such species would have had very skewed proportions of presences to absences and thus the standard incidence function models would not have been applicable. The issue of low sample size is further hampered by the use of binomial GLMs, which require an approximately equal proportion of ones and zeros (i.e. presences and absences). As the predicted pattern is reduced incidence in small fragments at the range edge relative to those in the range core the problem is particularly acute for my analyses, and forced me to reduce the number of target species from six to four.

Second, I was only able to collect snap shot presence-absence data for a single breeding season in each landscape. The use of single-year snap shot data implicitly assumes system equilibrium. However, incidence function model parameters have been found to vary between years for a given species (e.g. Moore & Hooper, 1975; Hinsley *et al.*, 1994, 1995) and it is possible that constructing incidence functions using time series data from my study systems may produce different results to the models constructed using the snap shot data (ter Braak *et al.*, 1997). For instance, Bellamy *et al.* (1996) found that the explanatory power of various environmental variables (including area) in explaining the presence-absence of various UK bird species varied between years, indicating temporal variability in the effect of area. That being said, other studies have reported minimal variation in incidence functions between years (e.g. Helzer & Jelinski, 1999), and as Hanski (2010; see also ter Braak *et al.*, 1997) states, "even a single snap shot has much information if presence-absence data are available for many islands." One proposed work around to deal with the issue of temporal variation in incidence functions is to sample many different landscapes (i.e. a space for time

substitution) and then fit mixed models with the landscape as a random variable (ter Braak *et al.,* 1997). This was not possible in my case as I only had data from four landscapes, which is too few to be used as a random variable in a mixed model (Zuur *et al.,* 2009).

Third, incidence functions have been found to be sensitive to other variables, such as the surrounding matrix type and the amount of surrounding forest (Hinsley *et al.,* 1996; Watson *et al.,* 2005; Mortelli *et al.,* 2010). Whilst I was able to keep matrix type constant (i.e. the main matrix was agricultural in all four landscapes) the amount of forest in the wider region did vary between landscapes. Species incidence has also been found to be affected by much finer scale landscape characteristics, such as the amount of hedgerows and microclimatic factors (Hinsley *et al.,* 1994, 1995; Mason, 2001). It would be impossible to select four landscapes in which all the factors affecting species incidence (see Whittaker & Fernández-Palacios, 2007 for a review) were kept equal, but the fact that species have been found to respond to fine scale local factors means it is hard to compare area incidence functions across landscapes with full confidence (Mason, 2001). These additional factors may explain the low pseudo R^2 values in many of my models. Further studies are required to determine to what extent area incidence function models are useful in the face of important local scale factors, such as microclimatic conditions. This issue is discussed further in Chapter 13.

12.6 Conclusions

Incidence function models are potentially useful as they are relatively easy to implement, require only a modest input of data, and generate results which are easily applicable to conservation in fragmented landscapes (Hanski, 1994a, b, 2010). The use of incidence function models also enables a greater focus on species specific responses to land use change; an approach which has been argued to have been overlooked in the decades since MacArthur & Wilson's (1967) equilibrium theory (Laurance, 2008; Hanski, 2010).

The applied benefit of incidence functions models depends on the extent to which the stability assumption (i.e. that the shape of a species' incidence function curve is constant across space and time) holds true. If species' incidence probabilities vary across space and time to a substantial degree, incidence function results from a certain time and location should not be used to instruct conservation actions in different contexts (Hinsley *et al.,* 1995; Watson *et al.,* 2005). My study was designed to assess the extent to which CIV values varied systematically across a species' range. In this regard the results were inconclusive, and for at least two of the four target species, the null hypothesis that there is no systematic variation in CIV values across the range cannot be rejected. For the remaining two species, the CIV values were lower in the range margin landscapes, but the models had poor explanatory power and the effect of the location and interaction terms in the models were rarely significant. That being said, whilst the hypothesised range effect was not conclusively proved, it is apparent that CIV values vary quite considerably across space for certain species (Table 12.2). My analyses also revealed additional limitations regarding the use of incidence function models in applied contexts, some of which have arguably been overlooked in previous studies. For example, my analyses revealed that it is possible to obtain sensible CIV values and incidence function plots even in cases where the predictor variables explain next to no variation in the response variable. As such, future incidence function studies should report the various model statistics (e.g. pseudo R^2, effect and significance of area) in addition to the standard plots and CIV values. In sum, according to my results the utility of incidence function models in generating useful conservation information is questionable.

CHAPTER 13: DISCUSSION

13.1 Conservation actions in patchy landscapes

The various empirical chapters (4-12) each contain specific discussion and conservation implication sections. My goal here is to draw out and summarise the general implications which arise from my analyses.

13.1.1 Strategic conservation planning

The conservation of fragmented landscapes has generally been based on theories derived from oceanic island biogeography (Diamond, 1975a; Whittaker *et al.*, 2005). However, as has been argued elsewhere in the thesis, habitat islands differ from true islands in numerous respects (Lindenmayer & Fischer, 2006; Watling & Donnelly, 2006; Laurance, 2008). Thus, a central aim of the thesis was to provide a comprehensive evaluation of macroecological patterns in purely habitat island systems. From this evaluation it is possible to derive information which can be used to inform protected area network design in habitat island systems.

Firstly, and unsurprisingly, ISARs were almost universally positive and thus at the most basic level a single large reserve is to be preferred over a single small reserve (Diamond, 1975a). Even at this basic level, the situation is not that simple as the predominant ISAR form was convex (e.g. the power model) and thus the rate of species accumulation with area decreases with increasing area. Thus, when incorporating constraints such as the cost of land, a smaller reserve may be more efficient than a larger one, above a certain area. However, such a simplistic question (i.e. the choice between a single large and a single small reserve) is rarely relevant to real world scenarios. Historically, the question, at least in academic circles, has been whether a single large (SL) reserve is preferable over several small (SS) reserves of the same area (i.e. the SLOSS debate). The SLOSS debate has its origins in the 1970s, the catalyst being provided by Diamond's (1975a) publication of a set of

reserve design criteria ostensibly derived from island biogeography theory. In a series of papers, Diamond and others argued that a single large reserve was the best strategy as, amongst other factors, many species have large minimum area requirements which cannot be met using small reserves (Terborgh, 1974; Diamond, 1976; Wilcox & Murphy, 1985). Wilcove *et al.* (1986) also championed the SL approach, arguing that small reserves will be dominated by edge habitats, with relatively little core habitat; the latter being known to be important for specialist species (Laurance *et al.,* 2002). Island biogeography theory also dictates the need for at least one large island to be present in the region in order to provide a source of immigrants to "rescue" populations that became extirpated from the smaller patches (Brown & Kodric-Brown, 1977; Wilcove *et al.,* 1986). Further studies in favour of the SL approach have argued that even if SS reserves contain more species than a SL reserve, many of these additional species are likely to be generalist species of little conservation concern (e.g. Terborgh, 1974; Whitcomb *et al.,* 1981). Interestingly, my Chapter 7 analyses seemingly back up this argument as I indeed found that smaller forest fragments contained a relatively high proportion of generalist species. A number of more recent studies have also provided evidence in support of the SL strategy. For example, Robinson (1999) found that a number of species have been lost from the BCI plot following isolation, and he concludes that alone small reserves such as BCI cannot sustain full levels of diversity.

Opponents of the SL strategy argued that a SS approach was better for two main reasons. First, there is reduced risk of disturbance wiping out species if the reserves are spread across a network of reserves (Simberloff & Abele, 1976, 1982). In this vein, it was argued that splitting up reserves into SS units also acts as a barrier against the spread of disease (Dobson & May, 1986). Second, a set of SS reserves spread out across a region can better represent the full set of habitats in the region, which in turn allows for a greater representation of specialised species (Simberloff & Abele, 1982).

The SLOSS debate has largely subsided, with the general agreement being that the answer to the SLOSS question is "it depends" (Whittaker & Fernández-Palacios, 2007). For example, in a comprehensive review of SLOSS studies, Ovaskainen (2002) found that there was basically a 50-50 split in terms of the main conclusions, i.e. half the studies recommended SL, and *vice versa*. As I showed in Chapter 8, the answer for a given system depends on the nestedness of the archipelago (see Fig. 8.3); with a number of smaller factors affecting the answer through secondary effects on nestedness (Ovaskainen, 2002). As I found no universal patterns regarding nestedness in habitat islands, it is thus not surprising that there is no universal answer to the SLOSS question. There has also been a realisation that whilst the SL strategy might be best for representing the most species, a number of species will only be represented in a protected area network through the inclusion of often small unique habitats/patches (Simberloff & Abele, 1982; Watson, 2003). My minimum set analyses in Chapter 8 support this argument as I found that to represent all species in the network it was generally necessary to include a high proportion of sites in the landscape in the network. Furthermore, in Chapter 6 I showed that, for my habitat island datasets, SACs constructed using island data were generally quite steep, at least relative to the ISAR (see Fig. 13.1 for an example using one of my habitat island datasets). The use of SACs in SLOSS type questions has a long history (see Quinn & Harrison, 1988), and steep SACs indicate that there is high turnover (or low nestedness) across islands and thus provide some support for the SS strategy.

In the end, the real answer to the reserve design question comes down to the amount of resources available, although what is actually meant by small and large reserves is also an important consideration. Very rarely is there a clean choice between a SL reserve or SS reserves; let alone the ability to decide exactly which fragments should be protected. However, it is important that conservation biogeography provides the theory and tools with

which managers and policy makers can apply to their landscapes (Ladle & Whittaker, 2011). In this regard, I am cautious to recommend either the SL or SS strategy based on my Chapter 6 & 8 results. Rather, I would argue that, depending on the data at hand, managers and conservation planners would be better served by avoiding making *a priori* assumptions. In cases where a set of habitat islands are to be destroyed, for example for logging concessions, and an opportunity for conserving a subset of islands is provided, it is likely that surveys will be conducted for at least certain taxa (e.g. see Watson *et al.*, 2009 for an example from Madagascar). With such survey data it is relatively simple to construct SACs, compute nestedness metrics, and run minimum set reserve planning algorithms (like that developed in Chapter 8), which may provide a coherent conservation strategy for the region at hand. In cases where there are no survey data, and no time to collect such data, the situation is obviously more problematic. In these cases, my results indicate that, without evidence to the contrary, systems should not be assumed to be nested *a priori*. A sensible strategy would then be to conserve the largest island possible (ideally the largest, but this is unlikely to be an available option in many instances) alongside a small set of smaller islands covering a range of major habitats (i.e. those habitats easily identifiable without site specific sampling data).

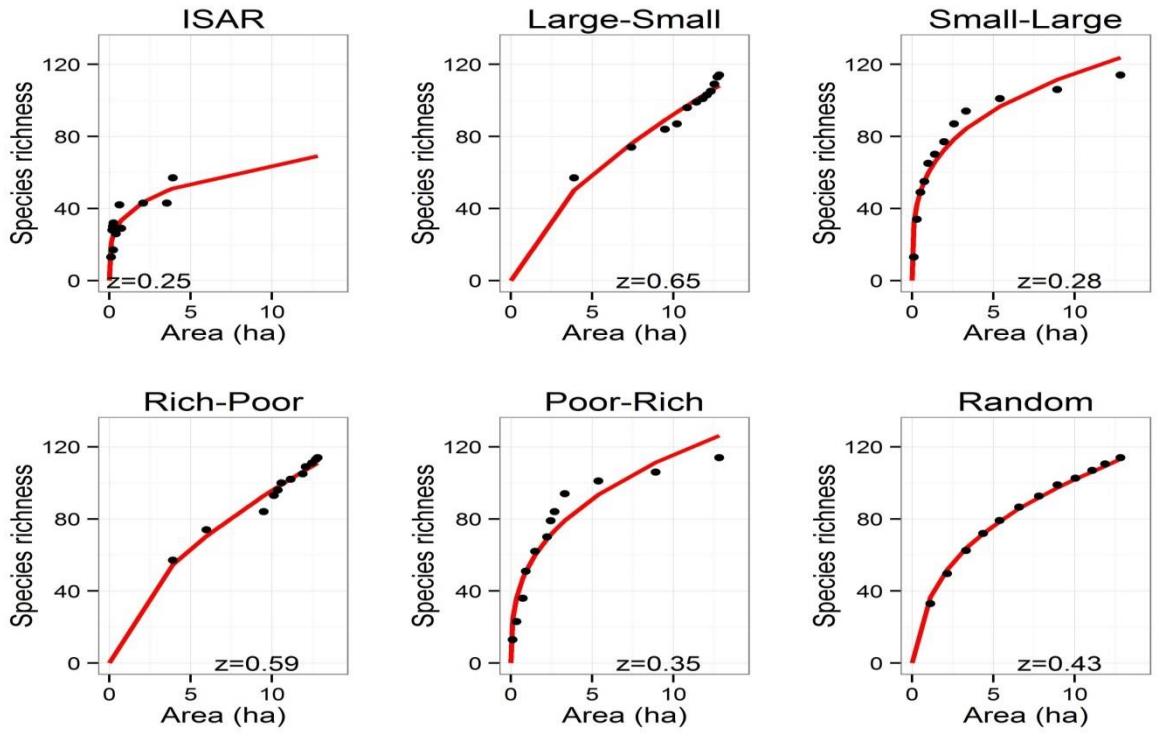


Figure 13.1 Exemplar fits of the power (non-linear) species–area relationship model to six different data structures, constructed from the same habitat island dataset: Gavish *et al.* (2012; invertebrates in a study system from Israel; number of islands = 12; number of species = 114). For each plot the individual habitat islands (black circles) are shown, alongside the fit of the power model (solid red line). The *z* value of the model is given in each instance. In the ISAR plot (top left), the fit of the power model has been extrapolated to the total cumulative area of all habitat islands in the system using the parameters derived from model fitting process; that is, the extrapolated area data point was not included in the model fitting process.

13.1.2 The assemblage deconstruction approach

Typically, empirical ecological and macroecological analyses are based on lists of all species encountered during a sampling exercise, e.g. all birds seen or heard in a patch of forest. However, it has become increasingly apparent that the aggregation in a sample of species with differences in key ecological properties, such as body size, dispersal ability or habitat affinity and specialisation can act to obscure patterns of interest for particular subsets of species; a theory I have termed the amalgamation hypothesis throughout this thesis. Thus, a number of recent studies have focused on splitting samples into various subsets prior to analysis and then exploring patterns of interest in each subset separately (e.g. Magurran &

Henderson, 2003; Ulrich & Zalewski, 2006; Bommarco *et al.,* 2010). For example, in Chapter 7, I have shown that the amalgamation of specialist and generalist bird species in forest fragment datasets can act to mask the loss of specialist species, generally the species of most conservation concern, in response to habitat insularisation (see also Bommarco *et al.,* 2010). To take another example, this time in the context of SADs, Labra *et al.* (2005) used a deconstruction approach to look at invasive and native species, finding that for US birds, on average, invasive species obtain higher maximum abundances than native species. Reportedly, this is due to the higher habitat generalisation of many invasive species, allowing them to reach higher abundances within a given community (Labra *et al.,* 2005). Furthermore, these different types of species are likely maintained and assembled through different processes and dynamics (Magurran & Henderson, 2003), and thus the identification of the important processes underpinning macroecological patterns may be challenging in such datasets.

Failure to take into account the effect on results of the amalgamation of different types of species in samples, for instance, in the above cases, generalist and invasive species, may result in inappropriate conservation plans being formulated (Matthews *et al.,* 2014a, c). Thus, future conservation biogeography studies should conduct analyses using various subsets of species, as well as using the whole assemblage. Admittedly, it is not always possible to classify species prior to data analysis. For example, in Chapter 7 it took considerable effort to classify the 1000 or so bird species into forest specialists and generalists, and there is more life history information available for birds relative to most other taxa (Gaston & Blackburn, 2000). But at the very least, studies need to consider how any patterns observed may be an artefact of the types of species included. Equally, the results of past conservation studies based on the total number of species sampled should be interpreted

with a degree of caution. A deconstruction approach for macroecology in general is advocated and described in depth by Marquet *et al*. (2004).

13.1.3 The impact of the matrix on ISAR *z* values

At a coarse scale, I observed that the *z* value of the ISAR is linked to the type of island, which in turn is linked to the matrix type in each dataset (Fig. 5.3). Whilst there was considerable scatter, the slope of the ISAR increased with decreasing matrix permeability, as previously predicted (e.g. Driscoll, 2005; Koh & Ghazoul, 2010; Pereira *et al.*, 2014). It follows that improving the quality and permeability of the matrix at fine scales will go some way to ameliorating the effects of habitat loss and fragmentation (Pereira & Daily, 2006). This area of research is the focus of countryside biogeography studies, which, amongst other things, have attempted to adapt traditional island theory in order for it to be more applicable to terrestrial systems and habitat islands (Daily *et al.*, 2003; Pereira & Daily, 2006; Koh *et al.*, 2010; Mendenhall *et al.*, 2014a, b). Countryside biogeography studies have shown that improvements to the matrix can benefit species persistence, as a more "ecologically friendly" matrix can (a) provide resources (e.g. prey) which species can utilise (Gascon *et al.*, 1999); (b) increase the ease of inter-patch dispersal, which in turn, reduces dispersal mortality, increases rescue effects, and reduces issues such as genetic drift in habitat islands (Ricketts, 2001; Haynes & Cronin, 2003; Joly *et al.*, 2003); and (c) supplement the habitat of the islands themselves (Brotons *et al.*, 2003). In regards to (c), studies have shown that even when all natural habitat is destroyed the number of species in the landscape does not drop to zero (as is the case in true island studies), as certain species can survive purely in the matrix habitat (Pereira & Daily, 2006). However, my Chapter 7 results indicate that these species are likely to be predominantly matrix generalist species, and thus the positive implications of this observation are questionable. I mentioned fine scale matrix improvements above, as coarse scale matrix changes are unlikely to be logistically possible, e.g. changing from an urban

matrix to secondary forest. Additionally, it is relatively easy to create policies for encouraging fine scale matrix improvements. For instance, European Union farmers are encouraged to maintain hedgerows and field margins as part of the Common Agricultural Policy (Stoate *et al.*, 2001).

13.1.4 Issues with incidence functions

My analyses revealed a number of issues with using standard incidence function models in ecological analyses, including the sensitivity of model parameters to (potentially) missing species data, and the fact that sensible plots can be constructed even when area explains very little variance in the response variable. Although I did not observe the hypothesised systematic variation in CIV values across the range, it was apparent that CIV values can vary quite substantially across space for certain species. For example, for robin, the CIV value in the French landscape was <0.01 ha, whilst the CIV value in the Norwegian landscape was 28.5 ha. This finding reinforces the arguments of previous authors who have called for protected areas and conservation plans to be focused in areas within a species range where the species has higher abundance and incidence (Gaston, 2003; Gaston & Rodrigues, 2003). Coupled with previous studies that have found that incidence function curves and CIV values can vary between years for the same landscape, and across landscapes due to e.g. matrix type (Hinsley *et al.*, 1994; Watson *et al.*, 2005), my analyses bring into question the utility of incidence function models (at least in studies which assume spatial and/or temporal stability) as conservation tools.

13.2 Are there any macroecological generalities in regards to patchy landscapes?

The previous section (13.1) was mainly focused on the lack of generality in habitat island systems. However, it is important, where possible, to identify broad general patterns around which future macroecological theory can be built (MacArthur & Wilson, 1967; Lawton,

1999; Gaston & Blackburn, 2000). In the subsequent sections, I attempt to identify such general patterns within my analyses.

13.2.1 Species abundance distributions

It is hard to derive general patterns regarding SADs in fragmented landscapes, due to the aforementioned amalgamation effect whereby the combination of different types of species (e.g. native & invasive) results in multimodal SADs, or at least skewed lognormal distributions (see Chapter 11). However, it is clear that the shape of the SAD is sensitive to the effects of habitat destruction and land use change. In chapter 10, I showed that the alpha parameter of the gambin model significantly decreased from native forest through plantation forest to agricultural pasture. This gradient is one of decreasing community complexity (i.e. a decline in the diversity of the community and the number of structuring processes) (Ugland *et al.,* 2007). Native forest arthropod assemblages exhibited lognormal type SADs, whilst pastureland assemblages typically exhibited more logseries-like SADs. These results align well with the conclusions of previous applied SAD studies (see Gray, 1987). The lognormal distribution has been proposed as a means of accurately modelling undisturbed communities and thus represents a measure by which deviation from such 'equilibrium' can be measured (e.g. May, 1975; Hill *et al.,* 1995). The suitability of the lognormal, as a general SAD model, derives from the Central Limit Theorem, in which a large number of biotic and abiotic factors affecting population sizes act multiplicatively to generate a lognormal distribution of abundances (May, 1975). In contrast, SADs in disturbed communities have been shown to follow distributions close to the logseries (e.g. Hill *et al.,* 1995). Early papers on the topic (e.g. Gray & Mirza, 1979; Gray *et al.,* 1979) focused on the effects of organic pollution on marine benthic communities, largely in Norway and Scotland, and used departure from lognormality to indicate the impact of pollution on community structure (but see Tokeshi 1993 for criticisms of their method). Following a mild pollution event, departure from

lognormality was argued to result, as a few species became more abundant whilst rare species remained at the same abundance or became extirpated from the system. Recent papers have provided further evidence of the utility of this approach (e.g. Tang *et al.,* 2010; Kim *et al.,* 2013). For example, Kim *et al.,* (2013) found that the SAD of macroinvertebrates in non-polluted streams (a key target group for ecosystem health) in Korea were best fitted by a lognormal model, whilst the SAD of macroinvertebrates in polluted streams was better fitted by a geometric model (a model with a similar curve shape to the logseries, in which the abundance of a particular species is proportional to the amount of limiting resource it has apportioned).

My results indicate that the SAD has considerable potential in applied habitat fragmentation research. Much of the recent work on SADs has placed a strong focus on the theoretical aspects of the pattern. This is a worthwhile and interesting endeavour, but arguably my results show that the SAD also has significant applied utility in numerous fields within ecology and biodiversity conservation. Study of the SAD allows inferences to be made regarding patterns in the commonness and rarity of species in a community; this being information of particular conservation and management importance. The SAD can act as an early warning for the effects of disturbance on ecological communities, as well as providing an intuitive tool for measuring the performance of ecological management prescriptions (Hågvar, 1994; Dornelas *et al.,* 2009). As the incidence and impact of disturbance is predicted to increase in the future (Sala *et al.,* 2000), the development and use of tools which are visually intuitive, easy to implement, are applicable in a broad variety of ecological contexts, and do not require substantial species specific data, is vital (Mouillot *et al.,* 2013). The results presented in Chapter 10 indicate that simple predictable changes in the shape of the SAD can be used to assess the impact of disturbance (e.g. land use change) on biodiversity in fragmented landscapes.

13.2.2 Species–area relationships

In my ISAR multimodel comparisons, I found that the power model was the best overall model according to all the model performance criteria used (i.e. generality, efficiency and number of best fits). It is reassuring that the power model consistently emerged as the best overall model given the preponderance of SAR studies that exclusively use this model (e.g. Watling & Donnelly, 2006). Our results are also consistent with those of Triantis *et al.* (2012) using true island datasets, in which it was found that the power model was the best-ranked model overall. Thus, whilst a multimodel inference approach is preferable, the power model is the most suitable choice of model for future ISAR studies if only one model is to be used.

In Chapter 5, I similarly found that a substantial degree of variation in the *z* and *c* values could be explained by various dataset characteristics. This implies that, to a certain degree, the form of the ISAR is a predictable phenomenon based on the qualities of the landscape under study. For example, ISAR *z* values were found to be affected by *Smin*, *Smax*, *Ascale* and *Lat* (see Chapter 5 for the descriptions of these variables). For *Lat* the effect was negative, meaning *z* decreased with increasing latitude (see also Connor & McCoy, 1979; Drakare *et al.*, 2006). This finding is consistent with the accepted general pattern of decreases in the size of species pools with increasing latitude, highlighted by the significant effect of *Smin* and *Smax* in our models; in turn, we argue that this points to an important role for climatic factors in driving the way in which species accumulate with area (Kalmar & Currie, 2006; Gerstner *et al.*, 2014). That *z* decreased with increasing *Ascale* seems to imply that the most dramatic increase in richness occurs over a low range of island areas and that beyond a certain size of island the gain in richness from additional area becomes slight. However, closer examination of the data revealed that *z* only decreased with *Ascale* for vertebrates (Fig. S4 in Appendix S3), seemingly indicating differences in the scale-dependency of the three taxa that were unapparent in comparison of the *z* values (discussed below). The negative

effect of *Smin* on the *z* value, coupled with the positive effect of *Smax*, makes sense as together these two variables represent the range in species richness within a dataset. If we assume that *Smin* typically occurs on a small island, while *Smax* occurs on one of the largest islands, then decreasing *Smin* and increasing *Smax* will necessarily result in a steeper slope in log–log space.

Whilst not a universal observation, it was found that for most datasets the slope of the ISAR curve for specialist species was steeper than that for generalist species. In terms of conservation-oriented ISAR applications, and based on the precautionary principle (Cooney, 2004), it seems sensible to assume this to be the case in all fragmented systems, in the absence of any evidence to the contrary. Nevertheless, my analyses focused exclusively on birds and it would be interesting to conduct similar examinations using different taxa (e.g. Magura *et al.*, 2001, 2008).

In general, I found that that the SAC was steeper than the ISAR constructed from the island data. This finding contrasts with the only other study to attempt this comparison, in which it was argued (with only one illustrative example) that the two curves should not differ much within a given dataset (Rosenzweig, 1995). In Chapter 6, I outlined a number of factors explaining this observation through decreasing the nestedness of the archipelago (e.g. multiple source pools, disturbance etc.). Each of these factors will vary in importance between habitat island systems and thus it is difficult to determine which factor is the most important. Nonetheless, in Chapter 7 I have shown that the relatively high incidence of generalist species in smaller fragments, which is a by-product of disturbance, reduces the slope of a multimodel ISAR curve for a number of habitat island datasets. This reduction in ISAR slope coupled with the reduced species overlap across islands provides an explanation for the steeper SAC relative to the ISAR. Regardless of the cause, this general finding leads to an important conclusion: that the ISAR generally under-predicts the number of species in

the archipelago, and by extension the richness of larger regions. This brings into question the extrapolation of ISARs to estimate the species richness of large areas; a common applied use of the ISAR (Whittaker *et al.*, 2005; Hui, 2008).

A lack of thresholds in the ISAR seems to be the norm in habitat islands (Chapter 4). This is the opposite finding from many landscape-scale threshold studies, which generally find a precipitous decline in richness after the amount of habitat in the landscape declines below a certain amount (see Andrén, 1994), highlighting that a particular pattern based on results from a different scale of analysis should not be assumed (Willis & Whittaker, 2002). In the modicum of datasets in which a threshold was detected, scrutiny of the piecewise models does not suggest that there is any clear tendency to a particular patch threshold size at which rate changes occur in the ISAR. This may reflect a general lack of consistency in response to area, variations in minimum area requirements of species, or it may arise from authors of my source papers having undertaken studies specifically to analyse the effects of one or more of an array of other variables (Maron *et al.*, 2012). Landscape features within studies are also important in determining breakpoints. For instance, variation in isolation and of matrix permeability can determine the existence of thresholds through the mitigation or accentuation of the effects of area loss *per se* (e.g. With & King, 1999; Brotons *et al.*, 2003; Watson *et al.*, 2005; Watling *et al.*, 2011). Another possible explanation of variation in breakpoint values is that the main two piecewise model forms (i.e. steep–shallow and shallow–steep) may represent different parts of a tri-phasic sigmoidal SAR in which different parts reflect different driving processes at different scales (cf. Rosenzweig, 1995; Lomolino & Weiser, 2001; Whittaker & Fernández-Palacios, 2007; Triantis *et al.*, 2012), but while this may apply to particular case studies, the present thesis does not provide a basis to claim generality of such a pattern. If a dataset only contains a few data points there are unlikely to be many points near the hypothetical breakpoint and there is thus potential to ‘miss’ the

threshold (cf. Lomolino & Weiser, 2001; Qie *et al.*, 2011). I found some support for this proposition, as piecewise models appeared to emerge mostly in datasets with a large number of islands. Notwithstanding these caveats, my results indicate that workers should not assume thresholds in habitat island ISARs *a priori*.

13.2.3 Nestedness

My search for general nestedness patterns in habitat islands might be considered a success or a failure, depending on how the results are interpreted. First, I found that datasets were rarely significantly compositionally nested when the PP algorithm was used as the null model. This lack of a significant pattern could be considered a generality of sorts. Furthermore, my nestedness analysis backs up recent studies, based predominantly on other types of island data (e.g. simulated and oceanic islands), that found significant nestedness to be relatively uncommon (e.g. Ulrich & Gotelli, 2007; Ulrich & Gotelli, 2013). This finding is noteworthy, as it contradicts the only previous analysis of multiple habitat island datasets (Watling & Donnelly, 2006), and reveals that conservation managers should not assume significant nestedness *a priori* in fragmented systems.

Second, when considering both the significantly nested datasets and all datasets, area seems to be an important environmental factor driving the observed nestedness in my datasets. This is conventionally taken as evidence for selective extinction being an important underlying mechanism (see Schouten *et al.*, 2007). This finding implies that extinction is a fairly deterministic process in these datasets, with the species possessing large area requirements being lost first. In regard to datasets that were significantly nested but for which the area ordered matrix was not correlated with the max matrix, it is likely that there are other important variables, such as disturbance and habitat heterogeneity, for which I lacked data (Fleishman & Murphy, 1999; Wang *et al.*, 2013). Incorporating functional traits into the nestedness analyses revealed useful information, but comparisons between the two forms of

nestedness analysis (Chapters 8-9) are problematic due to the fact that different null models were used in each; this was due in part to the different ways in which NODF and treeNODF work (Melo *et al.*, 2014).

13.3 The differences and similarities between true islands and habitat islands

Traditionally, habitat islands have been studied through the lens of island biogeographic theories that were derived and addressed empirically, mainly in the context of real islands (e.g. MacArthur & Wilson, 1963, 1967). During the 1970s, and following seminal studies investigating the applicability of island theory to terrestrial systems (e.g. Diamond, 1975a), there was a paradigm shift, so that studies began to focus on applying the island biogeographic theories to analyse terrestrial habitat islands (Whittaker *et al.*, 2005). However, significant academic debate surrounds this application, and several authors (e.g. Laurance, 2008) have criticised the potential of classical island theory to predict extinctions and species relaxation in fragmented systems. The major weakness with this approach is that it ignores the ineluctable observation that terrestrial habitat patches are affected by factors other than the main island biogeography variables of area and isolation (Cook *et al.*, 2002; Driscoll, 2005).

The most commonly discussed difference between true and habitat islands is the surrounding matrix. Generally speaking, habitat islands are surrounded by more permeable matrices than true islands, which in theory should result in easier inter-patch dispersal for most taxa (Ladle & Whittaker, 2011). Furthermore, many species can utilise resources in the matrix, increasing the effective size of habitat islands (Brotons *et al.*, 2003). As such, it has been postulated that habitat island archipelagos should have smaller *z* values than true island archipelagos due to these matrix effects (Rosenzweig, 1995). In my Chapter 5 analyses, I found this to be the case when simply comparing average *z* values across all true and habitat island datasets (see Fig. 13.2). However, this is a broad simplification and I observed great

variation in *z* values between habitat island types, according to the matrix. For example, urban habitat islands had average *z* values larger than the average for true islands. Thus, in this case it is likely that other characteristics of urban island systems (e.g. high disturbance) act to steepen the slope of the ISAR (Burgess & Sharpe, 1981).

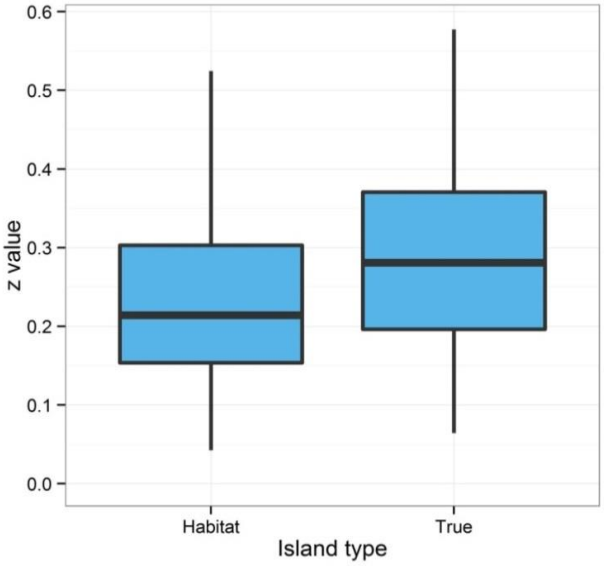


Figure 13.2 Variation in the *z* parameter of the power (log–log) model across habitat islands and true islands. Only *z* values from datasets with significant *z* values ($P < 0.05$) were included; for habitat island datasets this equated to 156 datasets, while there were 460 true island datasets. True islands comprised 125 oceanic island, 58 inland water-body, and 277 continental shelf island datasets. The box plots display the median (thick black line), the first and third quartiles (thin black box), and the minimum and maximum values (whiskers).

A further difference between oceanic islands and habitat islands, and one which is perhaps less discussed than matrix effects, relates to the dominant assembly processes in the two types of islands. Oceanic island biotas are generally assembled through a combination of speciation and immigration; the exact ratio depending on the distance of the archipelago from the mainland (MacArthur & Wilson, 1967). In contrast, habitat islands are predominantly immigration driven systems, regardless of the degree of isolation (i.e. speciation is rare; Rosenzweig, 1995). This is an important distinction, and one which must be borne in mind when interpreting the results of my analyses in the context of conservation recommendations.

Habitat islands also differ from oceanic islands in that the former generally start with a full complement of species following isolation, and subsequently relax to a lower 'equilibrium' number of species (Tilman *et al.*, 1994; Halley *et al.*, 2014). In contrast, true oceanic islands emerge from the sea and start with zero species (Whittaker & Fernández-

Palacios, 2007). In this respect, habitat islands are more similar to continental shelf islands (or at least those that are also land-bridge islands), which also "start" with a full complement of species (Harris, 1984; Whittaker & Fernández-Palacios, 2007). Based on this analogy, it has been predicted that continental shelf islands should have intermediate *z* values between oceanic and habitat island systems, and, based on average *z* values, we found this be the case (Fig. 5.3). Obviously, there are exceptions to this pattern. For example, mountaintop islands have different dynamics from other habitat island systems and in some respects are more akin to oceanic islands (Harris, 1984), with climatic changes resulting in the contraction or expansion of mountaintop islands being likened to isotactic sea level change affecting oceanic islands (Heaney, 2000).

A final important distinction between the two island types is that true islands are generally thought of in reference to a continental pool of potential immigrants (Whittaker & Fernández-Palacios, 2007). In contrast, a number of habitat island systems do not have a set source pool, i.e. the islands in the archipelago are the only source, or islands draw species from multiple places in the landscape (i.e. there is no easily delineable source pool)(Harris, 1984). This is important, because the lack of a single source pool adds a degree of stochasticity to the system (Hubbell, 2001), which can reduce the strength of particular macroecological patterns. Multiple source pools are also problematic when explaining macroecological patterns on some oceanic archipelagos (Price & Clague, 2002), but the effects are likely accentuated in habitat islands. For example, the multiple source pools effect is possibly one of the reasons I did not observe significant nestedness in most of my datasets: if islands draw species from multiple places, the likelihood is that the system will not be nested even if, for example, selective extinction is occurring. This coupled with the disturbed nature of fragmented landscapes means it is possibly foolhardy to even attempt to identify macroecological rules which apply in all systems (Lawton, 1999).

In sum, there are both similarities and differences between habitat islands and true islands, but on the basis of my results it is evident that habitat islands should be considered distinct systems, with my results providing support for those arguing for the development of habitat island specific theory (e.g. Crooks *et al.*, 2001; Laurance, 2008).

13.4 Key limitations

13.4.1 Limitations regarding the synthetic analyses

One of the issues with any ecological synthetic or meta-analytical study is how to ensure that the sampling methodology chosen and the level of sampling effort are appropriate for the study system under focus, within a potential source paper. The datasets I was interested in collecting were counts of species on habitat islands of varying area (herein, SC studies). Almost all SC studies rely on sampling (i.e. total counts of individuals are very rare in the empirical literature) and thus it is a given that there will be some degree of uncertainty surrounding the data gathered (see Cressie *et al.,* 2009). Thus, meta-analyses and synthetic analyses based on published SC data are also affected by sampling issues.

There are two main potential data quality issues in SC studies, and by extension SC meta-analyses. First, an inappropriate sampling method may result in data that generate or contain spurious patterns. Second, variables other than area (e.g. disturbance) are not controlled for. In regards to the first problem, the literature is replete with examples of how sampling effort can influence the form of various macroecological patterns (e.g. Connor & McCoy, 1979; Green & Plotkin, 2007). For example, many species abundance distribution models are based on a particular sampling theory (e.g. Etienne, 2005). The exact influence of low sampling effort on the SAR is less well studied, though it can be surmised that low-intensity sampling is more likely to under-represent the richness of larger islands due to the relatively greater sampling effort needed. As such, the right hand side of the SAR curve may

be artificially lowered, favouring particular models over others. One way of addressing this issue may be to only include datasets from studies in which the collection of island species lists is the primary methodological aim of the study. By this means, one can be certain that the original authors considered carefully how the data were to be collected, and that their methods were peer-reviewed as broadly being fit-for-purpose. This is not always the case in ecological meta-analyses. For example, Whittaker (2010) argued that many datasets used in productivity–richness meta-analyses were generated as a by-product of the analyses of the source paper, i.e. the generation of these data was not the primary aim of the study. In such cases, the data are much more likely to contain problems. Thus, in my analyses I attempted to ensure that I did not include data that were generated as a by-product of sampling designed for other purposes. Additionally, where possible, I tried to determine whether studies used proportional sampling or not. For example, Watson *et al.* (2009; one of my datasets) used a proportional sampling methodology in combination with rarefaction methods to determine when sampling effort was sufficient. Unfortunately, such information is not always clearly provided in source papers. Whilst it is possible that others attempting to replicate my analyses may make different decisions on data suitability, I believe that this is not likely to affect more than a handful of the decisions on dataset screening, and that the results presented here should therefore be immune to this particular potential source of bias.

The second data quality problem in SC studies has regard to the Chapter 4 discussion that confounding variables can influence the performance of certain piecewise regression ISAR models (Matthews *et al.*, 2014b). However, this issue is hard to eliminate, as very few habitat island systems contain no confounding variables; by their very nature they are disturbed systems and thus, for instance, disturbance and isolation will vary between islands within a landscape. It is also the case that many potential confounding variables are not presented in source papers, and/or were not measured to begin with, prohibiting a systematic

analysis of a given set of key variables of potential interest. Partly, this is due to the fact that there are a very large number of potential confounding variables in any study system, and workers have limited resources and have framed their research with slightly differing goals. My approach to this problem was therefore to avoid including studies in which a confounding variable (e.g. isolation) was the primary variable of interest, i.e. over and above area. Furthermore, I have no reason to believe that the inclusion of confounding variables would be biased towards any particular taxon or island type.

One further issue relates to how I have attempted to elucidate variations in different macroecological patterns across taxa. I have generally adopted a deconstruction approach, whereby I split the datasets into three broad taxon categories (i.e. vertebrates, plants and invertebrates) and then analysed patterns within each of these taxa individually. Additionally, I have used this three-way classification as a variable in a set of GLMs. Generally speaking this approach has failed to identify interesting taxon-specific patterns (e.g. see Fig. 5.3). Whilst there is a possibly that no such patterns exist, my approach suffers from the problem that these three categories are simply too coarse. There is a huge variation within each category, in terms of life history, dispersal ability and scale of perception. For example, vertebrates can be split into mammals, fish, reptiles and amphibians. Even within birds, there are substantial differences between how, for example, a small finch and a large hawk will respond to variations in landscape pattern (Wiens & Milne, 1989; Dunford & Freemark, 2005). The same intra-taxon variation is apparent within invertebrates and plants (e.g. Thomas, 2000; Henle *et al.,* 2004). However, such coarse groupings were necessary in order to avoid samples sizes too small to enable accurate statistical inference. Thus, future research interested in these types of macroecological questions should focus on finer-scale taxonomic delineation. One possible approach would be to focus on a specific taxon (e.g. mammals, Lepidoptera), rather than sourcing data on all taxa, and then deconstruct this taxon based on,

for example, dispersal ability and trophic status, and then re-analyse the data. Such an approach has been conducted before, but generally only for birds (e.g. Şekercioğlu *et al.*, 2002; Newbold *et al.*, 2013; Bregman *et al.*, 2014).

13.4.2 Limitations regarding my fieldwork data

There are four limitations regarding my fieldwork data. First, due to time constraints, I was only able to sample each landscape over a single breeding season (i.e. repeated sampling within a single year). The primary aim of this data collection was to construct incidence function (IF) models, and ideally I would have been able to fit (a) IF models which account for imperfect detection (Mackenzie *et al.*, 2002, 2006), and (b) the 'robust' hierarchical Bayesian IF models of Royle & Kéry (2007) and Risk *et al.* (2011). These various IF models deal with imperfect species detectability through the use of sample data that span multiple years. Additionally, data from multiple years allow better estimation of metapopulation IF model parameters (Oksanen, 2004; Risk *et al.*, 2011), though, it is still possible to derive useful parameter estimates from single year data (Oksanen, 2004). Perhaps the biggest potential issue is that I am unable to determine how robust the IF patterns are to inter-annual variation, i.e. how much the IF curves vary from year to year, for a given species. Previous studies have observed variation in species IFs between years (e.g. Hinsley *et al.*, 1994), and I cannot fully discount the possibility that the metapopulation type dynamics of my target species vary across years (e.g. Royle & Kéry, 2007).

Second, I have not sampled all fragments within each landscape, and it is not clear to what extent the inclusion of additional sites might change the results (Risk *et al.*, 2011). This issue has arisen primarily for logistical reasons, i.e. there is a trade-off between the number of fragments sampled, and the number of times each fragment can be sampled (Sutherland, 2006; Cressie *et al.*, 2009). I decided it was necessary to sample each fragment multiple times, resulting in fewer fragments sampled, in order to account for imperfect detection in my

models. Almost all fragmentation studies face this quandary and very few studies are able to sample all fragments in a landscape. Furthermore, how does one define a landscape? This is a conceptual question which has plagued landscape ecologists for decades (Naveh & Lieberman, 1984; Lindenmayer & Fischer, 2006), and none of my four fieldwork regions were located in neatly delineated "landscapes". Thus, even if I had had the resources to sample more fragments, I would have had to make an arbitrary decision to cease sampling somewhere. In sum, whilst it is possible that these additional fragments may have altered my results, I am confident that my reasonably large sample size (i.e. approximately 40 fragments in each landscape) allows me to make robust inferences. For comparison, the median number of habitat islands sampled in the 229 datasets I sourced from the literature was only 14.

Third, I have only sampled two range core and two range edge landscapes. Again, this was due to logistical reasons. Ideally, to answer this type of question one would sample landscapes along a gradient from the range core to the range edge (Gaston, 1994), and observe changes in species IFs along this gradient. Notwithstanding this fact, clear patterns have emerged, at least for some species, using only the four above landscapes. Furthermore, these landscapes accounted for roughly 150 fragments (each of which I sampled two-three times). To sample 40 fragments multiple times (and across multiple years) at more than four points along a range gradient would require substantial resources.

Fourth, whilst care was taken to select landscapes that were comparable in terms of matrix type, land use history and amount of forest in the surrounding landscape, it was not possible, for several reasons, to select landscapes that were identical. First, land use practices vary across countries. For example, agricultural practices in northern France and southern Spain are quite different, with the former involving intensive machine-based agriculture, and the latter involving more traditional pasture and grazing-based farming. Second, the amount of deforestation also varies across the continent. For instance, whilst the Norwegian

landscape was located in a region considered the most intensively farmed landscape in the country, most of Scandinavia is relatively well forested compared to central Europe and the UK. However, if these landscape differences have introduced any bias into the results, it would make my conclusions more conservative, because the most "hostile" landscape was one of the core sites: northern France. Furthermore, landscape connectivity and the amount of surrounding habitat were also lower in the French landscape, and were highest in the Norwegian landscape. Thus, it would be expected that incidence patterns would be negatively affected in this core landscapes relative to the edge landscapes, but this was not the pattern observed. There is also the possibility that differences in fragment structure between landscapes might result in varying species detectability across landscapes. For example, the worst case scenario in my analysis would be that the two range edge landscapes had much more dense forest than the two core landscapes, and thus my target species were less likely to be recorded in the edge sites. In my opinion this was not the case, and there was in fact a greater degree of variation in vegetation density within landscapes rather than between landscapes. Moreover, the target species were generally recorded in most large fragments regardless of the landscape; it was the smaller fragments where species were more likely to be absent. The smaller fragments were much easier to sample in all four landscapes and generally had less understorey vegetation. Furthermore, the target species were not cryptic species and were easily identifiable by call (e.g. wood pigeon, blue tit). In all landscapes the majority of identifications were based on calls, and the forest was never so dense that a call would not carry for 50m. Arguably, as with most ecological samples, the complete species counts for each fragment are an underestimate, but as these were not used in a comparative sense, this is less of a problem.

13.5 Future research opportunities

13.5.1 General research lacunae

I have identified six main avenues for future research to build on the results and conclusions presented in the thesis. The first involves a finer-scale analysis of the variation in macroecological patterns between different taxa in habitat island systems (section 13.4.1). Second, a more mechanistic approach to habitat island research is needed, because the identification of macroecological patterns, whilst interesting, is of little practical value without the identification of the processes driving the patterns (Gaston & Blackburn, 2000; McGill, 2003; Whittaker *et al.,* 2005). Whilst I have attempted to examine mechanisms where possible, I have been constrained by the lack of available data. In particular, there is a need to collect more environmental data from fragmented landscapes alongside species lists. This lack of environmental data partly motivated me to collect data on a number of additional variables in my fieldwork sites. Such data are useful because they allow one to examine macroecological patterns in relation to other variables. For example, environmental data can be integrated with the NODF and treeNODF nestedness metrics to evaluate patterns in the nestedness of species (Almeida-Neto *et al.,* 2008; Melo *et al.,* 2014). An interesting problem to investigate in this regard is the role of isolation in driving macroecological patterns in habitat islands. In most of my analyses (e.g. Chapter 9) isolation was generally unimportant. However, as noted by previous authors (e.g. Wang *et al.,* 2011), many of the patchy landscapes studied by ecologists are not effectively isolated for certain taxa. For example, inter-patch gaps of 100m are not much of a barrier to many bird species (Paradis *et al.,* 1998; Laurance *et al.,* 2002). Thus, an especially useful approach would be to create, possibly through experimental landscape manipulation (e.g. as per Laurance *et al.,* 2002), a number of fragments of identical area but with varying isolation. After sufficient time had passed it would then be possible to observe the point at which isolation becomes an important driver of

community assembly in patchy landscapes. A similar debate has been on-going in the oceanic island literature for many years (MacArthur & Wilson, 1967; Whittaker & Fernández-Palacios, 2007). Again, this question must be approached from the perspective of a particular taxon, as perception of inter-patch distances and thus isolation varies considerably between taxa (Wiens & Milne, 1989). Whilst costly, the experimental manipulation of landscapes is really the only method by which the various proposed drivers of species distributions in fragmented landscapes (e.g. isolation, disturbance, and shape) can be fully characterised. In "natural" (i.e. non-experimental) landscapes there are numerous confounding variables, which make summarising the role of individual variables problematic (e.g. Chapter 4).

Third, there is an opportunity for a more comprehensive linking of the findings of conservation biogeographic and macroecological studies focused on habitat islands, with on-the-ground conservation. Obviously, the aims and conclusions of certain macroecological studies are focused more towards large scale conservation, such as strategic conservation planning and protected area network design (Watson *et al.*, 2011). However, certain studies produce findings, recommendations, and ideas for future studies which are of relevance to more local scale conservation in fragmented landscapes. It is simply that the link between the two is not as apparent as in other ecological disciplines, e.g. population biology (e.g. Gilpin & Soulé, 1986). For example, in Matthews & Whittaker (2014a; summarised in Chapter 2) we show how the SAD can be of use in applied ecological contexts. Whilst the SAD is an interesting theoretical pattern in its own right (Fisher *et al.*, 1943; Williams, 1964; McGill *et al.*, 2007), and is useful for broad scale conservation measures (Pearce & Ferrier 2001; Dunstan *et al.* 2012), a number of SAD models can provide conservation information for more local scale applications. For instance, I showed in Chapter 10 how the alpha parameter of the gambin SAD model could be used as a metric to analyse the impact of land use change on arthropod assemblages. Whilst the mathematics and theory behind many SAD models is

relatively complex, this should not be a barrier preventing managers from using these models to provide useful information. Thus, future research should aim to further bridge this link between conservation macroecological and biogeographical studies, and information and/or tools for local scale conservation managers.

The following three future research opportunities are more methodological in nature, and are described in a standalone section.

13.5.2 Evaluating additional methodological approaches

Throughout the thesis I have adopted a number of different methodological approaches in order to examine macroecological and biogeographical patterns in habitat islands. Whilst most of these methodological approaches have been employed and described in the primary analyses and thesis papers, several have not. These represent methods which I have experimented with, but have not for one reason or another used in the primary analyses. These three methodological approaches have substantial potential for future macroecological and habitat island research. In this section, I outline these three methods, using examples based on the data collected during the thesis. The purpose of the below discussion is to illustrate the benefits of these approaches and provide recommendations for future research.

13.5.2.1 Multimodel inference and ISAR extrapolation

Common practice in SAR studies is to simply fit a single model to the data, generally the power model. This ignores model selection uncertainty. As noted by Burnham & Anderson (2002), such uncertainty is greater for extrapolations outside the range of the observed data, and models which provide good statistical fits to the observed data can provide very different predictions when extrapolated outside the range of the data. A more sensible approach to extrapolating a single model, then, is to calculate an averaged prediction of multiple models (i.e. multimodel inference; MMI), based for example on AIC_c weights (Burnham & Anderson, 2002).

The ISAR is often used for extrapolation purposes to, for example, estimate richness in larger areas, provide information for conservation planning frameworks and to predict the numbers of extinctions resulting from habitat loss (e.g. Sala *et al.,* 2005; van Vuuren *et al.*, 2006; Guilhaumon *et al.,* 2008; De Camargo & Currie, 2014). However, traditionally, only the power model (log–log) has been used (see Guilhaumon *et al.,* 2008). In Chapter 5, we found (a) that the power model does not always provide the best fit to empirical data, and (b) that there is generally a substantial deal of uncertainty surrounding the best model for a given dataset (e.g. the AIC_c weights for the fits of the ISAR models in Chapter 5 indicate that the weight of the best fitting model is rarely above 0.5), and (c) that the form, and thus predictions, of the various other ISAR models vary substantially (e.g. see Fig. 13.3). Thus, an MMI approach is arguably a more robust method for future ISAR extrapolation exercises (Guilhaumon *et al.,* 2008; Benchimol & Peres, 2013; Mazel *et al.,* 2014).

It is possible to use around 20 ISAR models in an ISAR extrapolation study (Chapter 5; Triantis *et al.,* 2012). However, it is generally not wise to include more models than data points in an analysis (Burnham & Anderson, 2002), and habitat island datasets generally have fewer than 20 data points. Thus, based on the results of previous analyses (Chapter 5; Triantis *et al.,* 2012) a subset of eight ISAR models, encompassing a range of ISAR forms, can be selected for extrapolation purposes: power, exponential, asymptotic, monod, Weibull3, Kobayashi, negative exponential and P2 (see Table 5.1 for model details). These models all performed relatively well based on the different model selection criteria (e.g. number of best fits, generality and efficiency) employed in Chapter 5. Thus, there is a sound *a priori* basis for using these models (Burnham & Anderson, 2002).

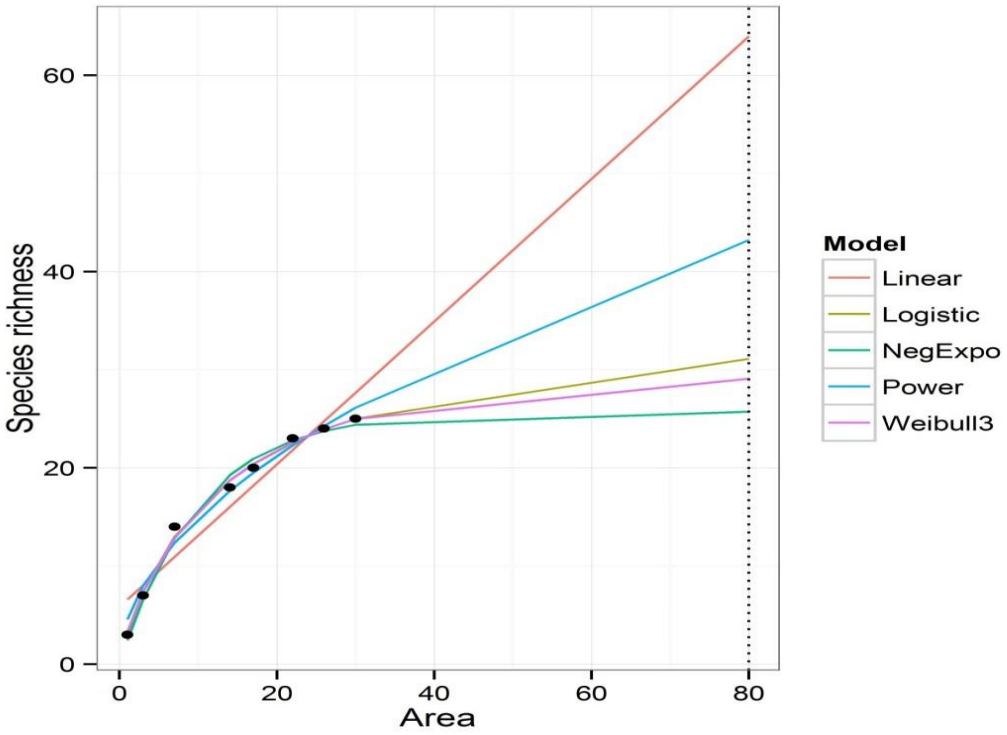


Figure 13.3 Model prediction uncertainty in ISAR extrapolation. Eight islands of varying size (1, 3, 7, 14, 17, 22, 26, and 30; undefined units) have been simulated with realistic numbers of species (3, 7, 14, 18, 20, 23, 24, 25). Five ISAR models (linear, logistic, negative exponential, power and Weibull3) were fitted to these eight data points. Using these model fits and parameter values the number of species on an island of size 80 (grey dotted line) was estimated for each model, and each curve has been extrapolated to its respective predicted value. Whilst the power model is the model most often used in such exercises, each of these five models has been found to provide the best fit to a particular dataset (Chapter 5).

13.5.2.1.1 Extrapolation methodology

To illustrate how an MMI approach can improve ISAR extrapolation predictions, I compared the extrapolation predictions of the power (log–log) model with the predictions of a multimodel ISAR curve using the aforementioned subset of eight ISAR models, and a multimodel ISAR curve using 17 models, for 150 habitat island datasets.

For each dataset, I selected a threshold cut-off value of 50%. I then removed the largest island, and all islands with area equal to or larger than 50% of the largest island (cf. Dengler, 2009). I have also experimented with other cut-off values (30, 70 & 90%). With this smaller set of islands, I then fitted the power model (log–log) to the data, and using the

parameters of the model fit I extrapolated the model to predict the species richness of the largest island in the dataset (i.e. the largest island pre-removal). I then fitted two multimodel ISAR curves (i.e. the eight models and 17 models) to the data, and repeated the extrapolation. The multimodel curves were constructed based on the AIC_c weights of each model fit (see the methods in Chapter 7). I then compared the predictions of the three models with the observed data (i.e. the observed richness of the largest island). A percentage accuracy measure (PM) was calculated which expressed the difference between the predicted and observed richness as a percentage. Finally, to test the precision of the predictions, I undertook a bootstrapping approach and created N bootstrap samples for each dataset using the residuals of the individual model fits (see Guilhaumon *et al.*, 2008). For each dataset and model fit, I created 500 bootstrap samples and fitted the three models to each, repeating the prediction exercise. From these bootstrap prediction values I then calculated 95% confidence intervals around the predictions of the three models.

13.5.2.1.2 Extrapolation results and conclusions

The multimodel ISAR curve constructed from the set of 17 ISAR models performed very poorly, and rarely provided the most accurate prediction. Thus, the following results are based on a direct comparison between the power (log–log) model (Pow), and the multimodel curve based on the subset of eight ISAR models (Sub).

In terms of raw species numbers, the Pow richness prediction was closest to the observed richness prediction for 49 datasets, whilst Sub was closer for 72 datasets. For 13 datasets the predictions were identical. Pow over-predicted the observed value (i.e. prediction > observed) 83 times, under-predicted 47 times, and was 'dead-on' (i.e. prediction = observed) four times. Sub over-predicted 68 times, under-predicted 59 times, and was dead-on (i.e. prediction = observed) seven times. In terms of the percentage accuracy measure (PM), the mean PM for Pow was 48.8% (2.5^{th} and 97.5^{th} quantiles = 0.56, 283.7) whilst the

mean PM for Sub was 28.0% (2.5th and 97.5th quantiles = 0.00, 101.56) (a lower value means a more accurate prediction). The confidence intervals (CI) of the predictions were also generally smaller for Sub, relative to Pow (mean CI width for Pow and Sub was 128.5 and 13.6, respectively.

Whilst acknowledging that the results of these analyses are preliminary, it appears that Sub provides a more general model for extrapolation of the ISAR than Pow. However, Pow did provide the most accurate prediction for a considerable minority of datasets; but as the CIs and PM of Sub were considerably smaller, Sub appears a much safer prediction tool. It is also apparent that the eight-model ISAR curve (i.e. a selection of models based on *a priori* science and the results of previous analyses) performs better than the 17-model curve (i.e. simply using all available models), providing further evidence of the importance of rigorous user decision making in model selection exercises (Burnham & Anderson, 2002). In sum, these results provide evidence for the potential of MMI approaches in ISAR extrapolation, and by extension predicting extinctions due to habitat loss. Future ISAR extrapolation studies should account for significant model selection uncertainty by constructing and extrapolating multimodel ISAR curves.

13.5.2.2 A Bayesian approach to incidence function modelling

Bayesian analysis involves the use of probability distributions to determine the level of uncertainty surrounding parameter values. Prior knowledge of potential parameter values is incorporated into the process and forms the 'prior distribution' (priors). If no prior knowledge exists, non-informative priors (i.e. a prior distribution which has minimal impact on the posterior distribution; generally Gaussian distributions with a mean of 0 and a large variance) can be employed. The posterior distribution represents the 'result' in Bayesian analysis and comprises a probability distribution of parameter values, and thus can be seen as more

informative than the single parameter value derived through more traditional methods (Bolker, 2008). The posterior distribution is derived through Bayes' theorem:

$$P(A \mid B) = (P(B \mid A) \cdot P(A)) \ / \ c$$

where $P(A \mid B)$ is the posterior probability, $P(B \mid A)$ is equivalent to the model likelihood (the probability of observing B, given A), P(A) is the prior distribution of parameter values, and *c* is a normalising constant which allows the posterior distribution to be a probability distribution. Posterior distributions are an "epistemological alternative to P-values" (Ellison, 2004, p.509) and represent a measure of the certainty that can be placed on parameter estimates. A number of point estimates relating to the posterior distribution are generally reported, including the mode and the mean of the posterior. The mean is generally preferred to the mode by proponents of Bayesian analysis as the former incorporates the entire distribution pattern, whilst the mode only reflects the peak of the posterior distribution (Bolker, 2008). In situations where a completely flat prior is used, the mode of the posterior will equal the maximum likelihood estimate.

Whereas frequentist approaches estimate the probability of observing the data, given the model/parameters ($P(B \mid A)$), Bayesian approaches provide a measure of the probability of the model/parameters being true, given the data ($P(A \mid B)$). A further difference is that frequentist approaches consider model parameters to be estimates of fixed true values, whereas Bayesian approaches consider them to be random variables.

The advantages of adopting a Bayesian approach are that the posterior probability distributions of parameter values allows for greater inference. Furthermore, a Bayesian view allows for decisions to be made on the plausibility of model parameter estimates, a process which is an important component of model testing (Etienne *et al.*, 2007). The downside is that most Bayesian approaches are analytically complex and generally require a greater amount of expert knowledge than information-theoretic approaches. Bayesian approaches

have been employed in a number of community ecology studies (see Ellison, 2004; Bolker, 2008), but are particularly popular in metapopulation and incidence-based studies (e.g. ter Braak & Etienne, 2003; Royle & Kéry, 2007; Risk *et al.*, 2011): a central topic of this thesis.

To highlight the advantages of Bayesian inference in incidence function based studies, I have provided two examples: 1) a Bayesian incidence function model, and 2) estimating the parameters of a standard logistic regression model using a Bayesian approach.

13.5.2.2.1 A Bayesian incidence function model

The metapopulation incidence function model (Mif; Hanski, 1997, 1999; Moilanen, 2002) models metapopulation dynamics as a function of area patch area and connectivity (see Chapter 2). Most standard applications of the Mif do not account for missing data or false absences. Depending on the study system and taxon, missing data and false absences may have significant impacts on the model parameters. Risk *et al.* (2011) developed a hierarchical Bayesian metapopulation incidence function model (termed the 'robust' model) which accounted for imperfect detection and missing data. It was shown that a Bayesian formulation of the Mif was advantageous as it provided a better idea of the uncertainty in the model parameters (i.e. the extinction and colonisation parameters; see Chapter 2) through analysis of the posterior distribution and the related credible intervals. The model was developed in the context of repeated sampling across years, and thus I was not able to use it in conjunction with my empirical data. Nonetheless, I wanted to show how the model could be useful for future research and have thus simulated an empirical dataset and fitted the model.

The simulated dataset consisted of the presence/absence of a single hypothetical species across 100 sites (of varying area and isolation, with values drawn from a lognormal distribution), and over 20 years (i.e. 20 presence/absence matrices were simulated). The probability that the species was present in a given site in a given year was taken from the Virginia rail (*Rallus limicola*) probability values presented in Risk *et al.* (2011). The

simulation process randomly removed the sample data of certain sites at certain years, and changed some presences to absences (see Risk *et al.,* 2011 for further details). The 'robust' model was fitted using the functions provided in Risk *et al.* (2011). Two chains were estimated for the model, and the number of iterations was set to 25,000 per chain. I did not set a burn-in for illustrative purposes, and retained one of every 50 iterations (i.e. thinning=50).

The results for the extinction parameter of the model (Mu) are given in Fig. 13.4. Fig.13.4a illustrates the trace of the first 500 iterations, and it is evident that for this parameter the model converges relatively quickly. This plot also illustrates the importance of using a burn-in (i.e. discarding the first n iterations), as there is a sharp decrease at the left hand side of Fig.13.4a. The posterior distribution of Mu is given in Fig.13.4b, and indicates a relatively sharp peak in the density. Thus, we can be relatively confident in the point estimate of the parameter, with the mean of the posterior distribution being 0.32 (though the long right hand tail likely present due to a lack of burn-in will have skewed the mean somewhat). Increasing Mu indicates an increased rate of extinction, and thus in this simulated system the probability of extinction is not particularly high.

I have provided this example to highlight how, if the data are available, more robust Mif models can be fitted than those used in Chapter 12. An interesting avenue for future research would be to expand the comparison of incidence function models in Chapter 12 to include hierarchical Bayesian models that incorporate missing data and imperfect sampling. This would enable a more thorough evaluation of different incidence models, and would provide vital information on the degree to which imperfect sampling affects incidence function model parameters (Moilanen, 2002; Mackenzie *et al.,* 2002; Risk *et al.,* 2011); models which can have important management implications (Mackenzie *et al.,* 2002; Royle & Kéry, 2007).

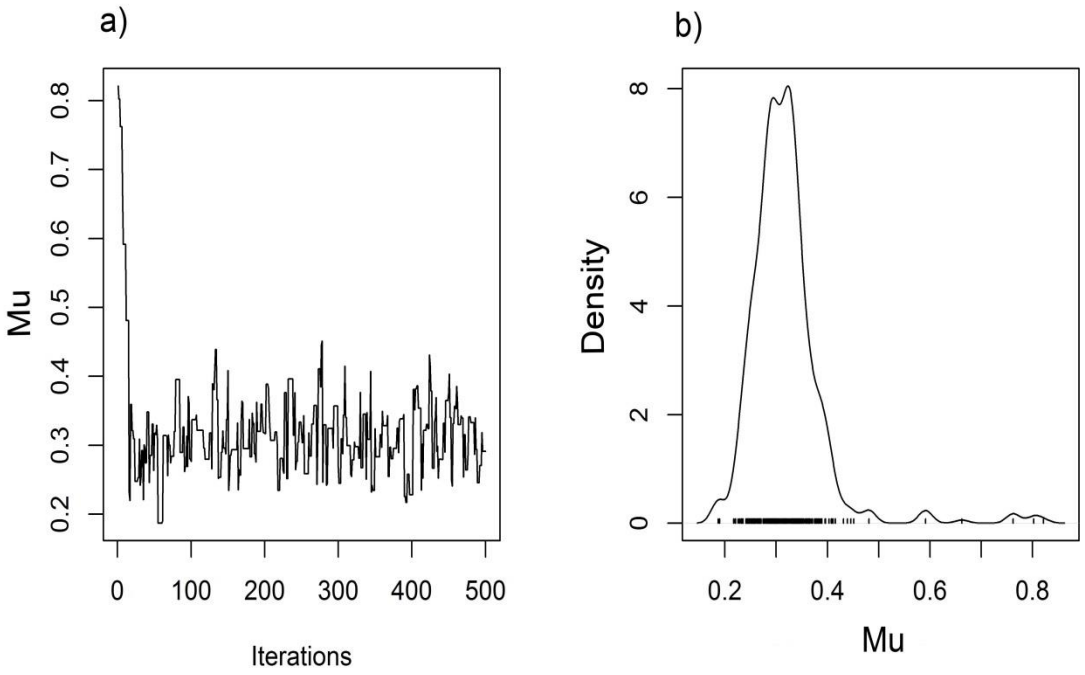


Figure 13.4 The results of a Bayesian metapopulation incidence function model fitted to a simulated dataset. The model parameter illustrated is the extinction parameter (Mu). a) represents the trace plot for the first 500 iterations of the model, and b) is the posterior distribution of the extinction parameter (Mu). The simulated data represent the presence/absence record of a single species across 100 sites of varying area and isolation, over 20 years. Two chains were estimated for the model, and the number of iterations was set to 25,000 per chain. I did not set a burn-in (for illustrative purposes), and have retained one of every 50 iterations.

13.5.2.2.2 Estimating regression parameters using a Bayesian approach

To provide a further example, I constructed a logistic regression model using my fieldwork data and estimated the parameters using two approaches: (a) the standard maximum likelihood (ML) method, and (b) a Bayesian approach. The response variable was the presence/absence of the short-toed tree creeper (*Certhia brachydactyla*) in 40 forest fragments of varying area in France. Two predictor variables were chosen based on initial data exploration: the average vertical vegetation density in a fragment (parameter a; see Chapter 3) and the area of each fragment (parameter b). To estimate the parameters using ML, I fitted a GLM of the binomial family (logit link) in R, storing the parameter estimates

for both predictors and the intercept. The model provided a reasonably good fit to the data, with the *P* value of a Pearson's X^2 goodness of fit test equal to 0.94.

I then estimated the parameters of the model using a Bayesian approach, and the WinBUGS software (Spiegelhalter *et al.*, 2003) and R2WinBUGS R package (Sturtz *et al.*, 2005). WinBUGS operates by sampling the posterior distributions using the Gibbs sampler. The logistic regression model was formulated in the WinBUGS language:

```
############################################################
##############WinBUGS script for logistic regression model############
#########################################################################
#st = presence-absence of the short-toed tree creeper; vd = vertical vegetation density of each
#fragment, area= area of each fragment.
model
{
for (i in 1:n) {
# Linear regression on logit
logit(p[i]) <- int + a*vd[i] + b*area[i]
# Likelihood function for each data point
st[i] ~ dbern(p[i])
}
int~dnorm(0,0.001) # Prior for intercept
a~dnorm(0,0.001) # Priors for other parameters
b~dnorm(0,0.001)
}
#########################################################################
```

Non-informative priors were selected for all three model parameters (i.e. ~ N (0, 1000)). Two chains were used, with 200,000 iterations per chain, and a burnin of 10,000. The posterior distributions of parameters a and b are given in Fig. 13.5. It can be seen that the Bayesian parameter point estimates (i.e. mean and mode of the posterior) differ slightly from the MLEs. These differences are relatively small, as is to be expected with a simple regression model and non-informative priors (though not completely uninformative), as used here. However, the figure neatly illustrates some of the benefits of using a Bayesian model fitting procedure as it can be seen that the posterior distributions provide a direct measure of the degree of belief that can be placed on the parameter estimates (Ellison, 2004), and by extension the uncertainty in these estimates. A frequentist confidence interval is often

interpreted as the probability that the true parameter lies within the confidence interval. However, the frequentist confidence interval is in fact the percentage of times the parameter of interest is expected fall within the interval in n runs of the experiment/samples from the true population. In contrast, the Bayesian credibility interval, defined as the interval in the centre of the distribution containing x% of the probability (Bolker, 2008), can actually be interpreted as the probability that a parameter lies within the interval (Ellison, 2004; Bolker, 2008). For example, in Fig.13.5a the 95% Bayesian credibility interval for parameter b is 1.1 - 4.5: quite a substantial range in terms of predicting changes in the response.

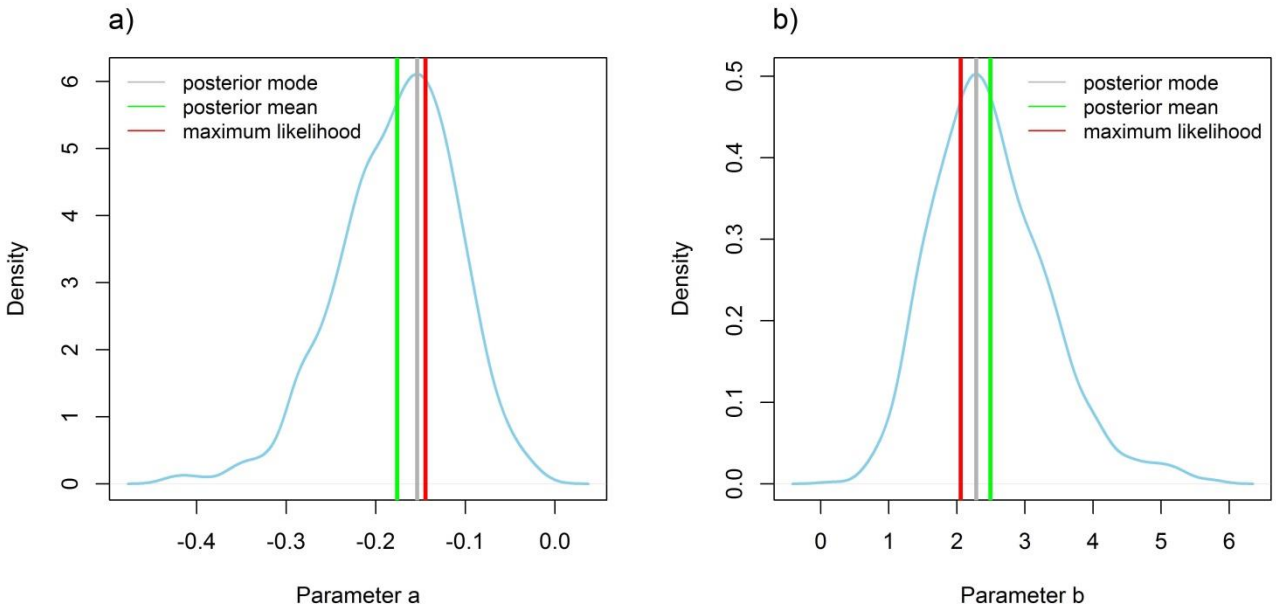


Figure 13.5 The posterior distributions of two logistic regression model parameters estimated using a Bayesian approach. The response variable was the presence/absence of the short-toed tree creeper (*Certhia brachydactyla*) in 40 forest fragments in France. In (a) the parameter 'a' refers to the effect of the mean vertical vegetation density per fragment on the response, and in (b) the parameter 'b' refers to the effect of the area of each fragment (log transformed). Non-informative priors were selected for all three model parameters (i.e. ~ N (0, 1000)). Two chains were used, with 200,000 iterations per chain, and a burnin of 10,000. In both (a) and (b), the mode and mean point estimates of the posterior are shown with solid coloured lines, as are the maximum likelihood estimates of the parameters calculated using a GLM of the binomial family.

13.5.2.3 Maximum entropy and SADs

The maximum entropy principle has recently been applied to the field of SADs (e.g. Banavar *et al.*, 2010; Harte, 2011; White *et al.*, 2012). Maximum entropy is a tool developed in the physical sciences and works by defining the most likely probability distribution as that which maximises entropy, subject to particular information constraints; the process starting with what we already know about the system and then filling out what we do not know by maximising entropy (Jaynes, 2003; Harte *et al.*, 2008; Frank, 2011). The method has been applied in the field of statistical mechanics to investigate how spheres (e.g. atoms) are distributed into boxes (energy levels, Banavar *et al.*, 2010). In SAD research, the spheres can represent individual organisms and the boxes can represent species. The theory behind the method is that multiple random factors that influence a particular pattern operate in different 'directions' and will ultimately cancel each other out in the aggregate (Frank, 2009, 2011), leaving an entirely random pattern with the exception of the restricting constraints. Once the user has defined the constraints, the least biased distribution can be determined via a constrained optimisation procedure, using the method of Lagrange multipliers (Jaynes, 2003). As such, it can be considered to minimise the information 'expressed in the final pattern' (Frank, 2011). For instance, Frank (2009) provides the example of the binomial distribution: if all the available information is minimised to the knowledge that a given observation is a set number of Bernoulli trials with an average number of successes, the observation will approximate the binomial distribution.

Opinions on the use of maximum entropy in SAD research are divided. Studies have shown that logseries and lognormal distributions, as well as neutral model SAD patterns, can be generated using a maximum-entropy method (Banavar & Maritan, 2007; Pueyo *et al.*, 2007; Harte, 2011; White *et al.*, 2012). For example, using four constraints, including the average per-species metabolic flux, and the average number of individuals per species, Harte

(2011) has shown that the SAD predicted by maximum entropy follows a logseries distribution. However, a number of workers have questioned the applicability of maximum entropy in ecology (e.g. Haegeman & Loreau, 2008; Haegeman & Etienne, 2010; He, 2010). Part of the problem with the use of maximum entropy in general is that, in order to obtain accurate results, the user must make a number of assumptions and select appropriate constraints; the user cannot identify missing constraints from the method itself. These choices are not trivial and have been shown to affect the outcome of the process (Haegeman & Etienne, 2010). Nonetheless, a recent application of the maximum entropy principle to over 15,000 ecological communities has shown that it is possible to characterise the majority of SAD structure when the only known information is the number of species and total number of individuals (White *et al.,* 2012; see also Pueyo *et al.,* 2007), thus largely circumventing the issue of constraint selection. The same study also proved that the parameter values of the logseries estimated by maximum entropy and by ML methods are equivalent (White *et al.,* 2012; see their supplementary material).

In my opinion, maximum entropy is a method with considerable potential for analysing probability distributions in macroecology, particularly SADs, and one which is ripe for further research and evaluation. In particular, there is a need to determine the consequences for particular methods of species missed during sampling (He, 2010), and a greater understanding is needed of how initial assumptions affect results (Haegeman & Etienne, 2010).

CHAPTER 14: CONCLUSIONS

The aim of this thesis was to determine whether there are any general rules regarding species diversity and species distributions in habitat islands, through the analysis of a wide range of habitat island datasets. As such, it is the first study to conduct such a thorough investigation of macroecological and biogeographic patterns in habitat islands. In the introduction, the question was posed whether it is possible to outline a well-specified macroecological theory of habitat islands? As with many questions in the natural sciences, the answer appears to be, yes, to some extent. Part of the problem with outlining a neat theory is that by their very nature habitat islands are generally disturbed systems, with numerous factors simultaneously affecting species diversity and species distributions. This type of issue plagues most areas of ecology (Lawton, 1999), but seems particularly acute in habitat island research. For example, in Chapter 4 it was apparent that the good fit of discontinuous piecewise models to a number of datasets was due to confounding variables.

Taken together, the results of my various analyses provide support to the various authors who have argued that, due to the high number of confounding variables in habitat island systems, the application of standard island biogeography theory is too simplistic (e.g. Laurance & Bierregaard, 1997; Ewers & Didham, 2006; Laurance, 2008). Nonetheless, it is possible to derive some general conclusions, namely, that: i) the power model is the best general ISAR model when considering habitat island datasets; ii) habitat island *z* values are generally lower than true island values, though certain habitat island systems have relatively high *z* values (e.g. islands set in an urban matrix); iii) there is little to evidence the existence of thresholds in the ISAR; iv) habitat islands are generally not significantly nested, though island area is important in determining the degree of nestedness (compositional and functional) observed; and v) the shape of the SAD in many habitat islands is influenced by the presence of numerous undesirable species. These general patterns, and more, are

discussed exhaustively in Chapter 13. The incidence function question remains equivocal to some degree. It does appear that there is some variation in species' incidence functions across space, even if the range effect is not as pronounced as hypothesised. Further research is required before we can be confident in using incidence function models as conservation tools.

What is also clear is that many of the macroecological and biogeographical patterns examined in this thesis are interlinked (Blackburn & Gaston, 2001). For example, the majority of habitat island datasets do not appear to be compositionally nested, underpinning the observed differences between the ISAR and the SAC, for a given dataset. Furthermore, part of the reason why these datasets are not nested is due to the large number of generalist and tourist species present in the smaller patches but not in many of the larger patches. The presence of these species also acts to depress the slope of the ISAR, and results in the multimodal SAD patterns observed.

It is also apparent that there are many differences between true and habitat islands (see section 13.3), not least that the issue of confounding variables appears to be more problematic in habitat islands. As such, it does not appear wise to continue to use the results of oceanic island studies to guide habitat island research and conservation.

My various analyses have shown that macroecological and biogeographical analyses can generate useful information for the conservation and management of biodiversity (Gaston & Blackburn, 2000; Whittaker *et al.*, 2005). For example, the results of Chapters 4-7, in combination with other recent studies (see Halley *et al.*, 2013), have enabled the development of a more robust and methodologically sound theory for using SARs in extrapolation exercises (e.g. to predict extinctions), albeit that admittedly further investigation is required. To take another example, at various points throughout the thesis I have highlighted how the SAD can be an important tool in applied ecology, providing information on the commonness and rarity of species overlooked when simple metrics such as species richness are used.

Looking forward, as reflected by the quote from Whittaker *et al.* (2005) in the introduction, what is needed is not so much the collection of more standard habitat island datasets (i.e. simple counts of the number of species across a set of islands), but, rather, the collection of datasets geared towards answering a particular hypothesis. Nevertheless, it is clear from the datasets that I collected (e.g. see Fig A3.1 in Appendix 3 for a map of the locations of the 207 datasets used in Chapter 5) that there is a gap in the global representation of habitat island datasets in Northern and Central Asia. Whilst there is no reason to presume that macroecological patterns in habitat islands in these locations will differ substantially from those in other locations, in the interest of completeness it would be useful to have access to datasets from these areas. Certain taxa (e.g. flowering plants) are also under-represented. Also, I believe that to truly advance the field of habitat island research we need to formulate novel research questions, and then design specific experiments or data collection protocols to answer these questions. For example, to answer the question of whether species incidence functions vary across a species range, it was necessary to design a specific sampling strategy involving the sampling of landscapes at various points across species' ranges. Particular research questions of interest include determining the effect of variables other than the standard biogeographic variables (i.e. area and isolation) on the form of the various macroecological patterns examined in this thesis, including different types of anthropogenic disturbance (e.g. hunting pressure). As stated above, most of my results' chapters include the caveat that the impact of such confounding variables on the pattern of interest is unknown. Thus, it would be interesting to design specific studies focused on these confounding variables.

Finally, an interesting future research avenue involves undertaking a more granular examination of the datasets already collected, e.g. by applying the assemblage deconstruction approach. I have shown at various points in this thesis that even adopting simple

classifications (e.g. native and invasive) can reveal interesting patterns. To achieve such outcomes it will be necessary to obtain substantially more species-specific trait data, particularly for non-vertebrate taxa. Thus, perhaps future fieldwork efforts would be better geared towards attempting to collect such trait data for habitat island species; even very basic data would allow us to begin separating the noise from the true macroecological order of nature.

REFERENCES

Abbott, I. (1983) The meaning of z in species/area regressions and the study of species turnover in island biogeography. *Oikos*, **41**, 385-390.

Aizen, M.A., Sabatino, M. & Tylianakis, J.M. (2012) Specialization and rarity predict nonrandom loss of interactions from mutualist networks. *Science*, **335**, 1486-1489.

Allan, B.F., Keesing, F. & Ostfeld, R.S. (2003) Effect of forest fragmentation on Lyme disease risk. *Conservation Biology*, **17**, 267-272.

Almeida-Neto, M. & Ulrich, W. (2010) A straightforward computational approach for quantifying nestedness using abundance data. *Environmental Modelling & Software*, **26**, 173-178.

Almeida-Neto, M., R. Guimarães Jr, P. & M. Lewinsohn, T. (2007) On nestedness analyses: rethinking matrix temperature and anti-nestedness. *Oikos*, **116**, 716-722.

Almeida-Neto, M., Guimarães, P., Guimarães, P.R., Loyola, R.D. & Ulrich, W. (2008) A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos*, **117**, 1227-1239.

Alonso, D., Ostling, A. & Etienne, R.S. (2008) The implicit assumption of symmetry and the species abundance distribution. *Ecology Letters*, **11**, 93-105.

Amarasekare, P. (1998) Allee effects in metapopulation dynamics. *The American Naturalist*, **152**, 298-302.

Anasiewicz, A. (1971) Observations on the bumble-bees in Lublin. *Ekol. Polska*, **19**, 401-417.

Anciães, M. & Marini, M.Â. (2000) The effects of fragmentation on fluctuating asymmetry in passerine birds of Brazilian tropical forests. *Journal of Applied Ecology*, **37**, 1013-1028.

Anderson, M.J. & Willis, T.J. (2003) Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. *Ecology*, **84**, 511-525.

Andrén, H. (1994) Effects of habitat fragmentation on birds and mammals in landscapes with different proportions of suitable habitat: a review. *Oikos*, **71**, 355-366.

Andrewartha, H.G. & Birch, L.C. (1954) *The distribution and abundance of animals*. University of Chicago Press, Chicago.

Aparicio, A. (2008) Descriptive analysis of the 'relictual' Mediterranean landscape in the Guadalquivir River valley (southern Spain): a baseline for scientific research and the development of conservation action plans. *Biodiversity and Conservation*, **17**, 2219-2232.

Aranda, S.C., Gabriel, R., Borges, P.A.V., Santos, A.M.C., Hortal, J., Baselga, A. & Lobo, J.M. (2013) How do different dispersal modes shape the species–area relationship? Evidence for between-group coherence in the Macaronesian flora. *Global Ecology and Biogeography*, **22**, 483-493.

Araújo, M.B. & Williams, P.H. (2000) Selecting areas for species persistence using occurrence data. *Biological Conservation*, **96**, 331-345.

Arrhenius, O. (1921) Species and area. *Journal of Ecology*, **9**, 95-99.

Asner, G.P., Elmore, A.J., Olander, L.P., Martin, R.E. & Harris, A.T. (2004) Grazing systems, ecosystem responses, and global change. *Annual Review of Environment and Resources*, **29**, 261-299.

Atmar, W. & Patterson, B.D. (1993) The measure of order and disorder in the distribution of species in fragmented habitat. *Oecologia*, **96**, 373-382.

Atmar, W. & Patterson, B.D. (1995) *The nestedness temperature calculator: a Visual Basic program*. AICS Research.

Azeria, E., Sanmartín, I., Ås, S., Carlson, A. & Burgess, N. (2007) Biogeographic patterns of the East African coastal forest vertebrate fauna. *Biodiversity and Conservation*, **16**, 883-912.

Azovsky, A.I. (2002) Size-dependent species-area relationships in benthos: is the world more diverse for microbes? *Ecography*, **25**, 273-282.

Baker, H., Stroud, D.A., Aebischer, N.J., Cranswick, P.A., Gregory, R.D., McSorley, C.A., Noble, D.G. & Rehfisch, M.M. (2006) Population estimates of birds in Great Britain and the United Kingdom. *British Birds*, **99**, 25-44.

Banavar, J.R. & Maritan, A. (2007) *The maximum relative entropy principle*. Available at: http://arxiv.org/abs/cond-mat/0703622 (accessed March 2014).

Banavar, J.R., Maritan, A. & Volkov, I. (2010) Applications of the principle of maximum entropy: from physics to ecology. *Journal of Physics: Condensed Matter*, **22**, 063101.

Banks-Leite, C., Ewers, R.M. & Metzger, J.P. (2012) Unraveling the drivers of community dissimilarity and species extinction in fragmented landscapes. *Ecology*, **93**, 2560-2569.

Barabás, G., D'Andrea, R., Rael, R., Meszéna, G. & Ostling, A. (2013) Emergent neutrality or hidden niches? *Oikos*, **122**, 1565-1572.

Barbosa, O. & Marquet, P.A. (2002) Effects of forest fragmentation on the beetle assemblage at the relict forest of Fray Jorge, Chile. *Oecologia*, **132**, 296-306.

Bartoń, K. (2012) *MuMIn: multi-model inference*.

Bascompte, J. & Sole, R.V. (1996) Habitat fragmentation and extinction thresholds in spatially explicit models. *Journal of Animal Ecology*, **65**, 465-473.

Baselga, A. (2012) The relationship between species replacement, dissimilarity derived from nestedness, and nestedness. *Global Ecology and Biogeography*, **21**, 1223-1232.

Baum, K.A., Haynes, K.J., Dillemuth, F.P. & Cronin, J.T. (2004) The matrix enhances the effectiveness of corridors and stepping stones. *Ecology*, **85**, 2671-2676.

Begon, M., Harper, J.L. & Townsend, C.R. (1990) *Ecology: individuals, populations and communities*, Second edn. Blackwell Scientific Publications, Boston, USA.

Bell, G. (2001) Neutral macroecology. *Science*, **293**, 2413-2418.

Bellamy, P.E., Hinsley, S.A. & Newton, I. (1996) Factors influencing bird species numbers in small woods in south-east England. *Journal of Applied Ecology*, **33**, 249-262.

Benchimol, M. & Peres, C.A. (2013) Anthropogenic modulators of species–area relationships in Neotropical primates: a continental-scale analysis of fragmented forest landscapes. *Diversity and Distributions*, **19**, 1339-1352.

Bender, D.J. & Fahrig, L. (2005) Matrix structure obscures the relationship between interpatch movement and patch size and isolation. *Ecology*, **86**, 1023-1033.

Bender, D.J., Contreras, T.A. & Fahrig, L. (1998) Habitat loss and population decline: a meta-analysis of the patch size effect *Ecology*, **79**, 517-533.

Bender, D.J., Tischendorf, L. & Fahrig, L. (2003) Using patch isolation metrics to predict animal movement in binary landscapes. *Landscape Ecology*, **18**, 17-39.

Bennett, J.M., Clarke, R.H., Thomson, J.R. & Mac Nally, R. (2014) Fragmentation, vegetation change and irruptive competitors affect recruitment of woodland birds. *Ecography*, doi: 10.1111/ecog.00936

Bentley, J.M. & Catterall, C.P. (1997) The use of bushland, corridors, and linear remnants by birds in south eastern Queensland, Australia. *Conservation Biology*, **11**, 1173-1189.

Betts, M.G., Forbes, G.J., Diamond, A.W. & Taylor, P.D. (2006) Independent effects of fragmentation on forest songbirds: an organism-based approach. *Ecological Applications*, **16**, 1076-1089.

Betts, M.G., Forbes, G.J. & Diamond, A.W. (2007) Thresholds in songbird occurrence in relation to landscape structure. *Conservation Biology*, **21**, 1046-1058.

Beyer, H.L. (2004) *Hawth's analysis tools for ArcGIS*. http://wwwspatialecologycom/htools.
Bibby, C.J., Burgess, N.D. & Hill, D.A. (2000) *Bird census techniques,* 2nd edn. Academic Press, London.
Bidwell, M.T., Green, A.J. & Clark, R.G. (2014) Random placement models predict species–area relationships in duck communities despite species aggregation. *Oikos*, **123**, 1499–1508.
Blackburn, T.M. & Gaston, K.J. (2001) Linking patterns in macroecology. *Journal of Animal Ecology*, **70**, 338-352.
Blackburn, T.M. & Gaston, K.J. (2002) Macroecology is distinct from biogeography. *Nature,* **418**, 723-723.
Blackburn, T.M. & Gaston, K.J. (eds.) (2003) *Macroecology: concepts and consequences: 43rd symposium of the British Ecological Society*. Cambridge University Press, Cambirdge.
Blackburn, T.M., Gaston, K.J., Quinn, R.M., Arnold, H. & Gregory, R.D. (1997) Of mice and wrens: the relation between abundance and geographic range size in British mammals and birds. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **352**, 419-427.
Blackburn, T.M., Gaston, K.J., Quinn, R.M. & Gregory, R.D. (1999) Do local abundances of British birds change with proximity to range edge? *Journal of Biogeography*, **26**, 493-505.
Blackburn, T.M., Petchey, O.L., Cassey, P. & Gaston, K.J. (2005) Functional diversity of mammalian predators and extinction in island birds. *Ecology*, **86**, 2916-2923.
Blake, J.G. (1991) Nested subsets and the distribution of birds on isolated woodlots. *Conservation Biology*, **5**, 58-66.
Bock, C.E. (1987) Distribution-abundance relationships of some Arizona landbirds: a matter of scale? *Ecology*, **68**, 124-129.
Boecklen, W.J. & Gotelli, N.J. (1984) Island biogeographic theory and conservation practice. *Biological Conservation*, **29**, 63-80.
Bolker, B.M. (2008) *Ecological models and data in R*. Princeton University Press, Princeton.
Bommarco, R., Biesmeijer, J.C., Meyer, B., Potts, S.G., Pöyry, J., Roberts, S.P.M., Steffan-Dewenter, I. & Öckinger, E. (2010) Dispersal capacity and diet breadth modify the response of wild bees to habitat loss. *Proceedings of the Royal Society B: Biological Sciences*, **277**, 2075-2082.
Bonte, D., Van Dyck, H., Bullock, J.M., Coulon, A., Delgado, M., Gibbs, M., Lehouck, V., Matthysen, E., Mustin, K., Saastamoinen, M., Schtickzelle, N., Stevens, V.M., Vandewoestijne, S., Baguette, M., Barton, K., Benton, T.G., Chaput-Bardy, A., Clobert, J., Dytham, C., Hovestadt, T., Meier, C.M., Palmer, S.C.F., Turlure, C. & Travis, J.M.J. (2012) Costs of dispersal. *Biological Reviews*, **87**, 290-312.
Borda-de-Água, L., Hubbell, S.P. & McAllister, M. (2002) Species-area curves, diversity indices, and species abundance distributions: a multifractal analysis. *The American Naturalist*, **159**, 138-155.
Borda-de-Água, L., Hubbell, S.P. & He, F. (2007) Scaling biodiversity under neutrality. *Scaling Biodiversity* (ed. by D. Storch, P.A. Marquet and J.H. Brown), pp. 347-375. Cambridge University Press, Cambridge.
Borda-de-Água, L., Borges, P.A.V., Hubbell, S.P. & Pereira, H.M. (2012) Spatial scaling of species abundance distributions. *Ecography*, **35**, 549-556.
Borges, P.A.V., Aguiar, C., Amaral, J., Amorim, I.R., André, G., Arraiol, A., A., B., Dinis, F., Enghoff, H., Gaspar, C., Ilharco, F., Mahnert, V., Melo, C., Pereira, F., Quartau, J.A., Ribeiro, S., Ribes, J., Serrano, A.R.M., Sousa, A.B., Strassen, R.Z., Vieira, L.,

Vieira, V., Vitorino, A. & Wunderlich, J. (2005) Ranking protected areas in the Azores using standardized sampling of soil epigean arthropods. *Biodiversity and Conservation*, **14**, 2029-2060.

Borges, P.A.V., Ugland, K.I., Dinis, F.O. & Gaspar, C.S. (2008) Insect and spider rarity in an oceanic island (Terceira, Azores): True rare and pseudo-rare species. *Insect Ecology and Conservation* (ed. by S. Fattorini), pp. 47-70. Research Signpost, Kerala.

Borges, P.A.V., Vieira, V., Amorim, I.R., Bicudo, N., Fritzén, N., Gaspar, C., Heleno, R., Hortal, J., Lissner, J., Logunov, D., Machado, A., Marcelino, J., Meijer, S.S., Melo, C., Mendonça, E.P., Moniz, J., Pereira, F., Santos, A.S., Simões, A.M. & Torrão, E. (2010) List of arthropods (Arthropoda). *A list of the terrestrial and marine biota from the Azores* (ed. by P.A.V. Borges, A. Costa, R. Cunha, R. Gabriel, V.A. Gonçalves, F. Martins, I. Melo, M. Parente, P. Raposeiro, P. Rodrigues, R.S. Santos, L. Silva, P. Vieira and V. Vieira), pp. 179-246. Princípia, Cascais.

Borregaard, M.K. & Rahbek, C. (2010) Causality of the relationship between geographic distribution and species abundance. *Quarterly Review of Biology*, **85**, 3-25.

Boyer, A.G. & Jetz, W. (2014) Extinctions and the loss of ecological function in island bird communities. *Global Ecology and Biogeography*, **23**, 679-688.

Bregman, T.P., Şekercioğlu, C.H. & Tobias, J.A. (2014) Global patterns and predictors of bird species responses to forest fragmentation: Implications for ecosystem function and conservation. *Biological Conservation*, **169**, 372-383.

Brook, B.W., Sodhi, N.S. & Ng, P.K.L. (2003) Catastrophic extinctions follow deforestation in Singapore. *Nature*, **424**, 420-426.

Brook, B.W., Traill, L.W. & Bradshaw, C.J.A. (2006) Minimum viable population sizes and global extinction risk are unrelated. *Ecology Letters*, **9**, 375-382.

Brooks, T. & Balmford, A. (1996) Atlantic forest extinctions. *Nature*, **380**, 115-115.

Brooks, T.M., Pimm, S.L. & Collar, N.J. (1997) Deforestation predicts the number of threatened birds in insular Southeast Asia. *Conservation Biology*, **11**, 382-394.

Brooks, T.M., Pimm, S.L. & Oyugi, J.O. (1999) Time lag between deforestation and bird extinction in tropical forest fragments. *Conservation Biology*, **13**, 1140-1150.

Brooks, T.M., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Rylands, A.B., Konstant, W.R., Flick, P., Pilgrim, J., Oldfield, S., Magin, G. & Hilton-Taylor, C. (2002) Habitat loss and extinction in the hotspots of biodiversity. *Conservation Biology*, **16**, 909-923.

Brooks, T.M., Brook, B.W., Koh, L.P., Pereira, H.M., Pimm, S.L., Rosenzweig, M.L. & Sodhi, N.S. (2011) Extinctions: consider all species. *Nature*, **474**, 284-284.

Brotons, L., Mönkkönen, M. & Martin, J.L. (2003) Are fragments islands? Landscape context and density-area relationships in boreal forest birds. *The American Naturalist*, **162**, 343-357.

Brown, J.H. (1971) Mammals on mountaintops: nonequilibrium insular biogeography. *The American Naturalist*, **105**, 467-478.

Brown, J.H. (1984) On the relationship between abundance and distribution of species. *The American Naturalist*, **124**, 255-279.

Brown, J.H. (1995) *Macroecology*. The University of Chicago Press, Chicago.

Brown, J.H. & Kodric-Brown, A. (1977) Turnover rates in insular biogeography: effect of immigration on extinction. *Ecology*, **58**, 445-449.

Brown, J.H. & Maurer, B.A. (1989) Macroecology: the division of food and space among species on continents. *Science* **243**, 1145-1150.

Brown, J.H., Mehlman, D.W. & Stevens, G.C. (1995) Spatial variation in abundance. *Ecology*, **76**, 2028-2043.

Brown, M. & Dinsmore, J.J. (1988) Habitat islands and the equilibrium theory of island biogeography: testing some predictions. *Oecologia*, **75**, 426-429.
Brualdi, R.A. & Sanderson, J.G. (1999) Nested species subsets, gaps, and discrepancy. *Oecologia*, **119**, 256-264.
Bulmer, M.G. (1974) On fitting the Poisson lognormal distribution to species-abundance data. *Biometrics*, **30**, 101-110.
Burgess, R.L. & Sharpe, D.M. (1981) *Forest island dynamics in man-dominated landscapes*. Springer-Verlag, New York.
Burnham, K.P. & Anderson, D.R. (2002) *Model selection and multi-model inference: a practical information-theoretic approach*, 2nd edn. Springer, New-York.
Burnham, K.P. & Anderson, D.R. (2004) Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods & Research*, **33**, 261-304.
Burns, K.C., Paul McHardy, R. & Pledger, S. (2009) The small-island effect: fact or artefact? *Ecography*, **32**, 269-276.
Cadotte, M., Albert, C.H. & Walker, S.C. (2013) The ecology of differences: assessing community assembly with trait and evolutionary distances. *Ecology Letters*, **16**, 1234-1244.
Cardoso, P., Lobo, J.M., Aranda, S.C., Dinis, F., Gaspar, C. & Borges, P.A.V. (2009) A spatial scale assessment of habitat effects on arthropod communities of an oceanic island. *Acta Oecologica*, **30**, 590-597.
Cardoso, P., Gaspar, C.S., Dinis, F. & Borges, P.A.V. (2010) Patterns of alpha and beta diversity of epigean arthropods at contrasting land-uses of an oceanic island (Terceira, Azores). *Terrestrial arthropods of Macronesia: biodiversity, ecology and evolution* (ed. by A.R.M. Serrano, P.a.V. Borges, M. Boieiro and P. Oromí), pp. 73–88. Sociedade Portuguesa de Entomologia, Lisbon.
Cardoso, P., Rigal, F., Fattorini, S., Terzopoulou, S. & Borges, P.A.V. (2013) Integrating landscape disturbance and indicator species in conservation studies. *PLOS ONE*, **8**, e63294.
Castelletta, M., Thiollay, J.-M. & Sodhi, N.S. (2005) The effects of extreme forest fragmentation on the bird community of Singapore Island. *Biological Conservation*, **121**, 135-155.
Catling, P. & Burt, R. (1995) Studies of the ground-dwelling mammals of eucalypt forests in south-eastern New South Wales: the effect of habitat variables on distribution and abundance. *Wildlife Research*, **22**, 271-288.
Channell, R. & Lomolino, M.V. (2000) Trajectories to extinction: spatial dynamics of the contraction of geographical ranges. *Journal of Biogeography*, **27**, 169-179.
Chase, J.M. & Knight, T.M. (2013) Scale-dependent effect sizes of ecological drivers on biodiversity: why standardised sampling is not enough. *Ecology Letters*, **16**, 17-26.
Chong, G.W. & Stohlgren, T.J. (2007) Species–area curves indicate the importance of habitats' contributions to regional biodiversity. *Ecological Indicators*, **7**, 387-395.
Claramunt, S., Derryberry, E.P., Remsen, J.V. & Brumfield, R.T. (2012) High dispersal ability inhibits speciation in a continental radiation of passerine birds. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 1567-1574.
Cochrane, M.A. & Laurance, W.F. (2002) Fire as a large-scale edge effect in Amazonian forests. *Journal of Tropical Ecology*, **18**, 311-325.
Coleman, B.D. (1981) On random placement and species-area relations. *Mathematical Biosciences*, **54**, 191-215.
Coleman, B.D., Mares, M.A., Willig, M.R. & Hsieh, Y.-H. (1982) Randomness, area, and species richness. *Ecology*, **63**, 1121-1133.

Condit, R. (1998) *Tropical forest census plots*. Springer-Verlag and R. G. Landes Company, Berlin and Georgetown, Texas.
Connolly, S.R. & Dornelas, M. (2011) Fitting and empirical evaluation of models for species abundance distributions. *Biological diversity: frontiers in measurement and assessment* (ed. by A.E. Magurran and B.J. Mcgill), pp. 123-140. Oxford University Press, Oxford.
Connolly, S.R., Hughes, T.P., Bellwood, D.R. & Karlson, R.H. (2005) Community structure of corals and reef fishes at multiple scales. *Science*, **309**, 1363-1365.
Connor, E.F. & McCoy, E.D. (1979) Statistics and biology of the species-area relationship. *The American Naturalist*, **113**, 791-833.
Cook, R.D. & Weisberg, S. (1982) *Residuals and influence in regression*. Chapman and Hall, London.
Cook, W.M., Lane, K.T., Foster, B.L. & Holt, R.D. (2002) Island theory, matrix effects and species richness patterns in habitat fragments. *Ecology Letters*, **5**, 619-623.
Cook, W.M., Anderson, R.M. & Schweiger, E.W. (2004) Is the matrix really inhospitable? Vole runway distribution in an experimentally fragmented landscape. *Oikos*, **104**, 5-14.
Cooney, R. (2004) Better safe than sorry? The precautionary principle and biodiversity conservation. *Oryx*, **38**, 357-358.
Corona, M., Vicente, A. & Novo, F. (1988) Long-term vegetation changes on the stabilized dunes of Doñana National Park (SW Spain). *Vegetatio*, **75**, 73-80.
Cramp, S. (1977) *Handbook of the birds of Europe, the Middle East, and North Africa: the birds of the Western Palearctic. Vol. 1, Ostrich to ducks*. Oxford University Press, Oxford.
Cramp, S. (ed.) (1985) *Handbook of the birds of Europe, the Middle East, and North Africa: the birds of the Western Palearctic. Vol. 4, Terns to woodpeckers.* Oxford University Press, Oxford.
Cramp, S. (ed.) (1988) *Handbook of the birds of Europe, the Middle East and North Africa: the Birds of the Western Palearctic. Vol. 5, Tyrant flycatchers to thrushes.* Oxford University Press, Oxford.
Cramp, S. & Brooks, D.J. (eds.) (1992) *Handbook of the Birds of Europe, the Middle East and North Africa: the birds of the Western Palearctic. Vol. 6, Warblers*. Oxford University Press, Oxford.
Cramp, S. & Perrins, C.M. (eds.) (1993) *Handbook of the Birds of Europe, the Middle East and North Africa:t he birds of the Western Palearctic. Vol. 7, Flycatchers to shrikes*. Oxford University Press, Oxford.
Cramp, S., Perrins, C.M. & Brooks, D.J. (eds.) (1994) *Handbook of the birds of Europe, the Middle East, and North Africa: the birds of the Western Palearctic. Vol. 8, Crows to finches*. Oxford University Press, Oxford.
Cressie, N., Calder, C.A., Clark, J.S., Hoef, J.M.V. & Wikle, C.K. (2009) Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. *Ecological Applications*, **19**, 553-570.
Crist, T.O. & Veech, J.A. (2006) Additive partitioning of rarefaction curves and species–area relationships: unifying α-, β- and γ-diversity with sample size and habitat area. *Ecology Letters*, **9**, 923-932.
Crooks, K.R. & Soulé, M.E. (1999) Mesopredator release and avifaunal extinctions in a fragmented system. *Nature*, **400**, 563-566.
Crooks, K.R., Suarez, A.V., Bolger, D.T. & Soulé, M.E. (2001) Extinction and colonization of birds on habitat islands. *Conservation Biology*, **15**, 159-172.

Crowe, T.M. (1979) Lots of weeds: insular phytogeography of vacant urban lots. *Journal of Biogeography*, **6**, 169-181.
Cutler, A.H. (1994) Nested biotas and biological conservation: metrics, mechanisms, and meaning of nestedness. *Landscape and Urban Planning*, **28**, 73-82.
Daily, G.C., Ceballos, G., Pacheco, J., Suzán, G. & Sánchez-Azofeifa, A. (2003) Countryside biogeography of Neotropical mammals: conservation opportunities in agricultural landscapes of Costa Rica. *Conservation Biology*, **17**, 1814-1826.
Dalecky, A., Chauvet, S., Ringuet, S., Claessens, O., Judas, J., Larue, M.A. & Cosson, J.-F. (2002) Large mammals on small islands: short term effects of forest fragmentation on the large mammal fauna in French Guiana. *Revue d'écologie*, **57**, 145-164.
Darlington, P.J. (1957) *Zoogeography: the geographical distribution of animals*. Wiley, New York.
Darwin, C. (1859) *On the origin of species by means of natural selection*. J. Murray, London.
Davis, B.N.K. & Jones, P.E. (1986) Insects on isolated colonies of common rock-rose Helianthemum chamaecistus. *Ecological Entomology*, **11**, 267-281.
De Camargo, R.X. & Currie, D.J. (2014) An empirical investigation of why species-area relationships overestimate species losses. *Ecology*, doi: 10.1890/13-2362.1
Debinski, D.M. & Holt, R.D. (2000) A survey and overview of fragmentation experiments. *Conservation Biology*, **14**, 342-355.
Dehling, D.M., Fritz, S.A., Töpfer, T., Päckert, M., Estler, P., Böhning-Gaese, K. & Schleuning, M. (2014) Functional and phylogenetic diversity and assemblage structure of frugivorous birds along an elevational gradient in the tropical Andes. *Ecography*, **37**, 1047–1055.
Del Re, A.C. (2010) *compute.es: Compute Effect Sizes.* R package version 0.2.
Dengler, J. (2009) Which function describes the species–area relationship best? A review and empirical evaluation. *Journal of Biogeography*, **36**, 728-744.
Dengler, J. (2010) Robust methods for detecting a small island effect. *Diversity and Distributions*, **16**, 256-266.
Dennis, R.L.H., Hardy, P.B. & Dapporto, L. (2012) Nestedness in island faunas: novel insights into island biogeography through butterfly community profiles of colonization ability and migration capacity. *Journal of Biogeography*, **39**, 1412-1426.
Desmet, P. & Cowling, R. (2004) Using the species-area relationship to set baseline targets for conservation. *Ecology and Society*, **9**, 11-33.
Devictor, V., Mouillot, D., Meynard, C., Jiguet, F., Thuiller, W. & Mouquet, N. (2010) Spatial mismatch and congruence between taxonomic, phylogenetic and functional diversity: the need for integrative conservation strategies in a changing world. *Ecology Letters*, **13**, 1030-1040.
Dewdney, A.K. (1998) A general theory of the sampling process with applications to the "veil line". *Theoretical Population Biology*, **54**, 294-302.
Diamond, J.M. (1975a) The island dilemma: lessons of modern biogeographic studies for the design of natural reserves. *Biological Conservation*, **7**, 129-146.
Diamond, J.M. (1975b) Assembly of species communities. *Ecology and evolution of communities* (ed. by M.L. Cody and J.M. Diamond), pp. 342-444. Harvard University Press, Cambridge, Massachusetts.
Diamond, J.M. (1976) Island biogeography and conservation: strategy and limitations. *Science*, **193**, 1027-1029.
Diamond, J.M. (1979) Community structure: is it random, or is it shaped by species differences and competition? *Population Dynamics* (ed. by R.M. Anderson, B.D. Turner and L.R. Taylor). Blackwell Scientific Publications, Oxford.

Diamond, J.M. (1984) Distributions of New Zealand birds on real and virtual islands. *New Zealand Journal of Ecology*, **7**, 37-55.
Diamond, J.M. & May, R.M. (1976) Island biogeography and the design of nature reserves. *Theoretical Ecology* (ed. by R.M. May), pp. 163–186. Blackwell Scientific Publications, Oxford.
Diamond, J.M., Bishop, K.D. & Balen, S.V. (1987) Bird survival in an isolated Javan woodland: island or mirror? *Conservation Biology*, **1**, 132-142.
Dick, C.W. (2001) Genetic rescue of remnant tropical trees by an alien pollinator. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **268**, 2391-2396.
Didham, R.K., Kapos, V. & Ewers, R.M. (2012) Rethinking the conceptual foundations of habitat fragmentation research. *Oikos*, **121**, 161-170.
Diehl, R.H. (2013) The airspace is habitat. *Trends in Ecology & Evolution*, **28**, 377-379.
Ding, Z., Feeley, K.J., Wang, Y., Pakeman, R.J. & Ding, P. (2013) Patterns of bird functional diversity on land-bridge island fragments. *Journal of Animal Ecology*, **82**, 781-790.
Diouf, K., Guilhaumon, F., Aliaume, C., Ndiaye, P., Chi, T.D. & Panfili, J. (2009) Effects of the environment on fish juvenile growth in West African stressful estuaries. *Estuarine, Coastal and Shelf Science*, **83**, 115-125.
Dobson, A., Lodge, D., Alder, J., Cumming, G.S., Keymer, J., McGlade, J., Mooney, H., Rusak, J.A., Sala, O., Wolters, V., Wall, D., Winfree, R. & Xenopoulos, M.A. (2006) Habitat loss, trophic collapse, and the decline of ecosystem services. *Ecology*, **87**, 1915-1924.
Dobson, A.P. & May, R.M. (1986) Disease and conservation. *Conservation biology: the science of scarcity and diversity* (ed. by M.E. Soulé), pp. 345-365. Sinauer Associates, Sunderland.
Dornelas, M. & Connolly, S.R. (2008) Multiple modes in a coral species abundance distribution. *Ecology Letters*, **11**, 1008-1016.
Dornelas, M., Moonen, A.C., Magurran, A.E. & Bàrberi, P. (2009) Species abundance distributions reveal environmental heterogeneity in modified landscapes. *Journal of Applied Ecology*, **46**, 666-672.
Dornelas, M., Soykan, C.U. & Ugland, K.I. (2011) Biodiversity and disturbance. *Biological diversity: frontiers in measurement and assessment* (ed. by A.E. Magurran and B.J. Mcgill), pp. 237-251. Oxford University Press, Oxford.
dos Anjos, L. (2001) Bird communities in five Atlantic forest fragments in southern Brazil. *Ornitologia Neotropical*, **12**, 11-27.
dos Anjos, L. & Boçon, R. (1999) Bird communities in natural forest patches in southern Brazil. *The Wilson Bulletin*, **111**, 397-414.
Drakare, S., Lennon, J.J. & Hillebrand, H. (2006) The imprint of the geographical, evolutionary and ecological context on species–area relationships. *Ecology Letters*, **9**, 215-227.
Dray, S. & Dufour, A.B. (2007) The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, **22**, 1-20.
Drinnan, I.N. (2005) The search for fragmentation thresholds in a Southern Sydney Suburb. *Biological Conservation*, **124**, 339-349.
Driscoll, D.A. (2005) Is the matrix a sea? Habitat specificity in a naturally fragmented landscape. *Ecological Entomology*, **30**, 8-16.
Driscoll, D.A., Banks, S.C., Barton, P.S., Lindenmayer, D.B. & Smith, A.L. (2013) Conceptual domain of the matrix in fragmented landscapes. *Trends in Ecology & Evolution*, **28,** 605-613.

Dunford, W. & Freemark, K. (2005) Matrix matters: effects of surrounding land uses on forest birds near Ottawa, Canada. *Landscape Ecology*, **20**, 497-511.
Dunstan, P.K., Bax, N.J., Foster, S.D., Williams, A. & Althaus, F. (2012) Identifying hotspots for biodiversity management using rank abundance distributions. *Diversity and Distributions*, **18**, 22-32.
Elith, J., Leathwick, J.R. & Hastie, T. (2008) A working guide to boosted regression trees. *Journal of Animal Ecology*, **77**, 802-813.
Ellison, A.M. (2004) Bayesian inference in ecology. *Ecology Letters*, **7**, 509-520.
Engen, S. (2001) A dynamic and spatial model with migration generating the log-Gaussian field of population densities. *Mathematical Biosciences*, **173**, 85-102.
ESRI (2011) *ArcGIS Desktop: Release 10*. Environmental Systems Research Institute.
Estades, C.F. (2001) The effect of breeding-habitat patch size on bird population density. *Landscape Ecology*, **16**, 161-173.
Etienne, R.S. (2005) A new sampling formula for neutral biodiversity. *Ecology Letters*, **8**, 253-260.
Etienne, R.S., Apol, M.E.F., Olff, H. & Weissing, F.J. (2007) Modes of speciation and the neutral theory of biodiversity. *Oikos*, **116**, 241-258.
Ewers, R.M. & Didham, R.K. (2006) Confounding factors in the detection of species responses to habitat fragmentation. *Biological Reviews*, **81**, 117-142.
Ewers, R.M. & Didham, R.K. (2007) The effect of fragment shape and species' sensitivity to habitat edges on animal population size. *Conservation Biology*, **21**, 926-936.
Ewers, R.M., Didham, R.K., Pearse, W.D., Lefebvre, V., Rosa, I.M.D., Carreiras, J.M.B., Lucas, R.M. & Reuman, D.C. (2013) Using landscape history to predict biodiversity patterns in fragmented landscapes. *Ecology Letters*, **16**, 1221-1233.
Fahrig, L. (1997) Relative effects of habitat loss and fragmentation on population extinction. *The Journal of Wildlife Management*, **61**, 603-610.
Fahrig, L. (1998) When does fragmentation of breeding habitat affect population survival? *Ecological Modelling*, **105**, 273-292.
Fahrig, L. (2001) How much habitat is enough? *Biological Conservation*, **100**, 65-74.
Fahrig, L. (2002) Effect of habitat fragmentation on the extinction threshold: a synthesis. *Ecological Applications*, **12**, 346-353.
Fahrig, L. (2003) Effects of habitat fragmentation on biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, **34**, 487-515.
Fahrig, L. (2013) Rethinking patch size and isolation effects: the habitat amount hypothesis. *Journal of Biogeography*, **40**, 1649-1663.
Fattorini, S. & Borges, P.A.V. (2012) Species-area relationships underestimate extinction rates. *Acta Oecologica*, **40**, 27-30.
Feeley, K. (2003) Analysis of avian communities in Lake Guri, Venezuela, using multiple assembly rule models. *Oecologia*, **137**, 104-113.
Ferraz, G., Russell, G.J., Stouffer, P.C., Bierregaard, R.O., Pimm, S.L. & Lovejoy, T.E. (2003) Rates of species loss from Amazonian forest fragments. *Proceedings of the National Academy of Sciences USA*, **100**, 14069-14073.
Fesl, C. (2002) Niche-oriented species-abundance models: different approaches of their application to larval Chironomid (Diptera) assemblages in a large river. *Journal of Animal Ecology*, **71**, 1085-1094.
Ficetola, G.F. & Denoël, M. (2009) Ecological thresholds: an assessment of methods to identify abrupt changes in species-habitat relationships. *Ecography*, **32**, 1075-1084.
Fischer, J. & Lindenmayer, D.B. (2002) Small patches can be valuable for biodiversity conservation: two case studies on birds in southeastern Australia. *Biological Conservation*, **106**, 129-136.

Fischer, J. & Lindenmayer, D.B. (2005a) Nestedness in fragmented landscapes: a case study on birds, arboreal marsupials and lizards. *Journal of Biogeography*, **32**, 1737-1750.
Fischer, J. & Lindenmayer, D.B. (2005b) Perfectly nested or significantly nested – an important difference for conservation management. *Oikos*, **109**, 485-494.
Fischer, J. & B. Lindenmayer, D. (2006) Beyond fragmentation: the continuum model for fauna research and conservation in human-modified landscapes. *Oikos*, **112**, 473-480.
Fischer, J. & Lindenmayer, D.B. (2007) Landscape modification and habitat fragmentation: a synthesis. *Global Ecology and Biogeography*, **16**, 265-280.
Fisher, R.A., Corbet, A.S. & Williams, C.B. (1943) The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology*, **12**, 42-58.
Fleishman, E. & Murphy, D.D. (1999) Patterns and processes of nestedness in a Great Basin butterfly community. *Oecologia*, **119**, 133-139.
Flynn, D.F.B., Gogol-Prokurat, M., Nogeire, T., Molinari, N., Richers, B.T., Lin, B.B., Simpson, N., Mayfield, M.M. & DeClerck, F. (2009) Loss of functional diversity under land use intensification across multiple taxa. *Ecology Letters*, **12**, 22-33.
Foggo, A., Attrill, M.J., Frost, M.T. & Rowden, A.A. (2003) Estimating marine species richness: an evaluation of six extrapolative techniques. *Marine Ecology Progress Series*, **248**, 15-26.
Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., Helkowski, J.H., Holloway, T., Howard, E.A., Kucharik, C.J., Monfreda, C., Patz, J.A., Prentice, I.C., Ramankutty, N. & Snyder, P.K. (2005) Global consequences of land use. *Science*, **309**, 570-574.
Foppen, R.P.B., Chardon, J.P. & Liefveld, W. (2000) Understanding the role of sink patches in source-sink metapopulations: reed warbler in an agricultural landscape. *Conservation Biology*, **14**, 1881-1892.
Ford, H.A. (1987) Bird communities on habitat islands in England. *Bird Study*, **34**, 205 - 218.
Ford, H.A. & Barrett, G. (1995) The role of birds and their conservation in agricultural systems. *People and nature conservation* (ed. by S.B.A. Bennett and T. Clark), pp. 123–128. Royal Zoologial Society of New South Wales, Mosman, Aust.
Forest Research Institute Malaysia (2013) *Pasoh species abundance*. Available at: http://www.ctfs.si.edu/site/Pasoh/abundance (accessed December 2013).
Fortuna, M.A., Krishna, A. & Bascompte, J. (2013) Habitat loss and the disassembly of mutalistic networks. *Oikos*, **122**, 938-942.
Frank, S.A. (2009) The common patterns of nature. *Journal of Evolutionary Biology*, **22**, 1563-1585.
Frank, S.A. (2011) Measurement scale in maximum entropy models of species abundance. *Journal of Evolutionary Biology*, **24**, 485-496.
Franklin, A.B., Noon, B.R. & George, T.L. (2002) What is habitat fragmentation? *Studies in Avian Biology*, **25**, 20-29.
Franklin, I.R. & Frankham, R. (1998) How large must populations be to retain evolutionary potential? *Animal Conservation*, **1**, 69-70.
Frishkoff, L.O., Karp, D.S., M'Gonigle, L.K., Mendenhall, C.D., Zook, J., Kremen, C., Hadly, E.A. & Daily, G.C. (2014) Loss of avian phylogenetic diversity in neotropical agricultural systems. *Science*, **345**, 1343-1346.
Fronhofer, E.A., Kubisch, A., Hilker, F.M., Hovestadt, T. & Poethke, H.J. (2012) Why are metapopulations so rare? *Ecology*, **93**, 1967-1978.
Galipaud, M., Gillingham, M.A.F., David, M. & Dechaume-Moncharmont, F.-X. (2014) Ecologists overestimate the importance of predictor variables in model averaging: a plea for cautious interpretations. *Methods in Ecology and Evolution*, **5**, 983–991.

Gascon, C., Lovejoy, T.E., Bierregaard Jr, R.O., Malcolm, J.R., Stouffer, P.C., Vasconcelos, H.L., Laurance, W.F., Zimmerman, B., Tocher, M. & Borges, S. (1999) Matrix habitat and species richness in tropical forest remnants. *Biological Conservation*, **91**, 223-229.

Gaspar, C., Borges, P.A.V. & Gaston, K.J. (2008) Diversity and distribution of arthropods in native forests of the Azores archipelago. *Arquipélago Life and marine Sciences*, **25**, 1-30.

Gaston, K.J. (1990) Patterns in the geographical ranges of species. *Biological Reviews*, **65**, 105-129.

Gaston, K.J. (1994) *Rarity*. Chapman & Hall, London.

Gaston, K.J. (2003) *The structure and dynamics of geographic ranges*. Oxford University Press, Oxford.

Gaston, K.J. & Lawton, J.H. (1990) Effects of scale and habitat on the relationship between regional distribution and local abundance. *Oikos*, **58**, 329-335.

Gaston, K.J. & Blackburn, T.M. (2000) *Patterns and process in macroecology*. Blackwell Science, Oxford.

Gaston, K.J. & Rodrigues, A.S.L. (2003) Reserve selection in regions with poor biological data. *Conservation Biology*, **17**, 188-195.

Gaston, K.J. & Blackburn, T.M. (2003) Dispersal and the interspecific abundance-occupancy relationship in British birds. *Global Ecology and Biogeography*, **12**, 373-379.

Gaston, K.J., Blackburn, T.M. & Gregory, R.D. (1997a) Interspecific abundance-range size relationships: range position and phylogeny. *Ecography*, **20**, 390-399.

Gaston, K.J., Blackburn, T.M. & Lawton, J.H. (1997b) Interspecific abundance-range size relationships: an appraisal of mechanisms. *Journal of Animal Ecology*, **66**, 579-601.

Gaston, K.J., Borges, P.A.V., He, F. & Gaspar, C. (2006) Abundance, spatial variance and occupancy: arthropod species distribution in the Azores. *Journal of Animal Ecology*, **75**, 646-656.

Gavish, Y., Ziv, Y. & Rosenzweig, M.L. (2012) Decoupling fragmentation from habitat loss for spiders in patchy agricultural landscapes. *Conservation Biology*, **26**, 150-159.

Gentile, G. & Argano, R. (2005) Island biogeography of the Mediterranean sea: the species–area relationship for terrestrial isopods. *Journal of Biogeography*, **32**, 1715-1726.

Gerstner, K., Dormann, C.F., Václavík, T., Kreft, H. & Seppelt, R. (2014) Accounting for geographical variation in species–area relationships improves the prediction of plant species richness at the global scale. *Journal of Biogeography*, **41**, 261-273.

Gibbons, D.W. & Gregory, R.D. (2006) Birds. *Ecological census techniques* (ed. by W.J. Sutherland), pp. 308-350. Cambridge University Press, Cambridge.

Gibson, L., Lynam, A.J., Bradshaw, C.J.A., He, F., Bickford, D.P., Woodruff, D.S., Bumrungsri, S. & Laurance, W.F. (2013) Near-complete extinction of native small mammal fauna 25 years after forest fragmentation. *Science*, **341**, 1508-1510.

Gilbert, B., Laurance, W.F., Leigh, E.G. & Nascimento, H.E.M. (2006) Can neutral theory predict the responses of amazonian tree communities to forest fragmentation? *The American Naturalist*, **168**, 304-317.

Gillespie, T.W. & Walter, H. (2001) Distribution of bird species richness at a regional scale in tropical dry forest of Central America. *Journal of Biogeography*, **28**, 651-662.

Gilpin, M.E. & Soulé, M.E. (1986) Minimum viable populations: processes of species and extinction. *Conservation Biology: The Science of Scarcity and Diversity* (ed. by M.E. Soulé). Sinauer Associates, Sunderland, Massachusetts.

Gleason, H.A. (1922) On the relation between species and area. *Ecology*, **3**, 158-162.

Gleason, H.A. (1926) The individualistic concept of the plant association. *Bulletin of the Torrey Botanical Club*, **53**, 7-26.

Gonzalez, A. & Chaneton, E.J. (2002) Heterotroph species extinction, abundance and biomass dynamics in an experimentally fragmented microecosystem. *Journal of Animal Ecology*, **71**, 594-602.
Gonzalez, A., Lawton, J.H., Gilbert, F.S., Blackburn, T.M. & Evans-Freke, I. (1998) Metapopulation dynamics, abundance, and distribution in a microecosystem. *Science*, **281**, 2045-2047.
Google Earth (2013) *Google Earth, version 7.1.2.2041*. http://www.google.com/earth/index.html.
Gotelli, N.J. (1991) Metapopulation models: the rescue effect, the propagule rain, and the core-satellite hypothesis. *The American Naturalist*, **138**, 768-776.
Gotelli, N.J. & Graves, G.R. (1990) Body size and the occurrence of avian species on land-bridge islands. *Journal of Biogeography*, **17**, 315-325.
Gotelli, N. & Entsminger, G. (2001a) Swap and fill algorithms in null model analysis: rethinking the knight's tour. *Oecologia*, **129**, 281-291.
Gotelli, N.J. & Entsminger, G.L. (2001b) *EcoSim: Null models software for ecology*. Acquired Intelligence Inc. & Kesey-Bear.
Gotelli, N.J. & Colwell, R.K. (2010) Estimating species richness. *Biological diversity: frontiers in measurement and assessment* (ed. by A.E. Magurran and B. Mcgill), pp. 39-54. Oxford University Press, Oxford.
Gotelli, N.J. & Ulrich, W. (2012) Statistical challenges in null model analysis. *Oikos*, **121**, 171-180.
Gould, S.J. (1979) An allometricinterpretation of species-area curves: the meaning of the coefficient. *The American Naturalist*, **114**, 335-343.
Grant, P.R. & Grant, B.R. (2006) Evolution of character displacement in Darwin's finches. *Science*, **313**, 224-226.
Gray, J.S. (1987) Species-abundance patterns. *Organization of Communities Past and Present* (ed. by J.H.R. Gee and P.S. Giller), pp. 53–68. Blackwell Science, Oxford.
Gray, J.S. & Mirza, F.B. (1979) A possible method for the detection of pollution-induced disturbance on marine benthic communities. *Marine Pollution Bulletin*, **10**, 142-146.
Gray, J.S., Waldichuk, M., Newton, A.J., Berry, R.J., Holden, A.V. & Pearson, T.H. (1979) Pollution-induced changes in populations. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, **286**, 545-561.
Gray, J.S., Ugland, K.I. & Lambshead, J. (2004a) On species accumulation and species–area curves. *Global Ecology and Biogeography*, **13**, 567-568.
Gray, J.S., Ugland, K.I. & Lambshead, J. (2004b) Species accumulation and species–area curves — a comment on Scheiner (2003). *Global Ecology and Biogeography*, **13**, 473-476.
Gray, J.S., Bjørgesæter, A. & Ugland, K.I. (2005) The impact of rare species on natural assemblages. *Journal of Animal Ecology*, **74**, 1131-1139.
Gray, J.S., Bjørgesæter, A. & Ugland, K.I. (2006a) On plotting species abundance distributions. *Journal of Animal Ecology*, **75**, 752-756.
Gray, J.S., Bjørgesæter, A., Ugland, K.I. & Frank, K. (2006b) Are there differences in structure between marine and terrestrial assemblages? *Journal of Experimental Marine Biology and Ecology*, **330**, 19-26.
Green, J.L. & Plotkin, J.B. (2007) A statistical theory for sampling species abundances. *Ecology Letters*, **10**, 1037-1045.
Gregory, R.D. & Gaston, K.J. (2000) Explanations of commonness and rarity in British breeding birds: separating resource use and resource availability. *Oikos*, **88**, 515-526.
Groffman, P., Baron, J., Blett, T., Gold, A., Goodman, I., Gunderson, L., Levinson, B., Palmer, M., Paerl, H., Peterson, G., Poff, N., Rejeski, D., Reynolds, J., Turner, M.,

Weathers, K. & Wiens, J. (2006) Ecological thresholds: the key to successful environmental management or an important concept with no practical application? *Ecosystems*, **9**, 1-13.
Grøtan, V. & Engen, S. (2009) *poilog: Poisson lognormal and bivariate Poisson lognormal distribution.* R package version 0.4.
Guardiola, M., Pino, J. & Rodà, F. (2013) Patch history and spatial scale modulate local plant extinction and extinction debt in habitat patches. *Diversity and Distributions*, **19**, 825-833.
Guilhaumon, F., Mouillot, D. & Gimenez, O. (2010) mmSAR: an R-package for multimodel species–area relationship inference. *Ecography*, **33**, 420-424.
Guilhaumon, F., Gimenez, O., Gaston, K.J. & Mouillot, D. (2008) Taxonomic and regional uncertainty in species-area relationships and the identification of richness hotspots. *Proceedings of the National Academy of Sciences USA*, **105**, 15458-15463.
Guo, Q., Brown, J.H., Valone & Thomas, J. (2000) Abundance and distribution of desert annuals: are spatial and temporal patterns related? *Journal of Ecology*, **88**, 551-560.
Haegeman, B. & Loreau, M. (2008) Limitations of entropy maximization in ecology. *Oikos*, **117**, 1700-1710.
Haegeman, B. & Etienne, R.S. (2010) Entropy maximization and the spatial distribution of species. *The American Naturalist*, **175**, E74-E90.
Hågvar, S. (1994) Log-normal distribution of dominance as an indicator of stressed soil microarthropod communities? *Acta Zoologica Fennica*, **195**, 71-80.
Haila, Y. (2002) A conceptual genealogy of fragmentation research: From island biogeography to landscape ecology. *Ecological Applications*, **12**, 321-334.
Halley, J.M., Sgardeli, V. & Monokrousos, N. (2013) Species–area relationships and extinction forecasts. *Annals of the New York Academy of Sciences*, **1286**, 50-61.
Halley, J.M., Sgardeli, V. & Triantis, K.A. (2014) Extinction debt and the species–area relationship: a neutral perspective. *Global Ecology and Biogeography*, **23**, 113-123.
Hanski, I. (1982) Dynamics of regional distribution: the core and satellite species hypothesis. *Oikos*, **38**, 210-221.
Hanski, I. (1992) Inferences from ecological incidence functions. *The American Naturalist*, **139**, 657-662.
Hanski, I. (1994a) Patch-occupancy dynamics in fragmented landscapes. *Trends in Ecology & Evolution*, **9**, 131-135.
Hanski, I. (1994b) A practical model of metapopulation dynamics. *Journal of Animal Ecology*, **63**, 151-162.
Hanski, I. (1997) Metapopulation dynamics: from concepts and observations to predictive models. *Metapopulation biology: ecology, genetics, and evolution* (ed. by I. Hanski and M.E. Gilpin), pp. 69-91. Academic Press, San Diego.
Hanski, I. (1998) Metapopulation dynamics. *Nature*, **396**, 41-49.
Hanski, I. (1999) *Metapopulation ecology*. Oxford University Press, Oxford.
Hanski, I. (2010) The theories of island biogeography and metapopulation dynamics. *The theory of island biogeography revisited* (ed. by J.B. Losos and R.E. Ricklefs), pp. 186–213. Princeton University Press, Princeton.
Hanski, I. & Gyllenberg, M. (1993) Two general metapopulation models and the core-satellite species hypothesis. *The American Naturalist*, **142**, 17-41.
Hanski, I. & Gyllenberg, M. (1997) Uniting two general patterns in the distribution of species. *Science*, **275**, 397-400.
Hanski, I., Kouki, J. & Halkka, A. (1993) Three explanations of the positive relationship between distribution and abundance of species. *Species diversity in ecological*

communities: historical and geographical perspectives (ed. by R.E. Ricklefs and D. Schluter), pp. 108-116. University of Chicago Press, Chicago.
Hanski, I., Moilanen, A., Pakkala, T. & Kuussaari, M. (1996) The quantitative incidence function model and persistence of an endangered butterfly metapopulation. *Conservation Biology*, **10**, 578-590.
Hanski, I., Zurita, G.A., Bellocq, M.I. & Rybicki, J. (2013) Species–fragmented area relationship. *Proceedings of the National Academy of Sciences USA*, **110**, 12715-12720.
Harris, L.D. (1984) *The fragmented forest: island biogeographic theory and the preservation of biotic diversity*. University of Chicago Press, Chicago.
Harte, J. (2011) *Maximum entropy and ecology: a theory of abundance, distribution, and energetics*. Oxford University Press.
Harte, J., Kinzig, A. & Green, J. (1999) Self-similarity in the distribution and abundance of species. *Science*, **284**, 334-336.
Harte, J., Zillio, T., Conlisk, E. & Smith, A.B. (2008) Maximum entropy and the state-variable approach to macroecology. *Ecology*, **89**, 2700-2711.
Haynes, K.J. & Cronin, J.T. (2003) Matrix composition affects the spatial ecology of a prairie planthopper. *Ecology*, **84**, 2856-2866.
He, F. (2010) Maximum entropy, logistic regression, and species abundance. *Oikos*, **119**, 578-582.
He, F. & Legendre, P. (1996) On species-area relations. *The American Naturalist*, **148**, 719-737.
He, F. & Hubbell, S.P. (2011) Species-area relationships always overestimate extinction rates from habitat loss. *Nature*, **473**, 368-371.
He, F. & Hubbell, S.P. (2012) He and Hubbell reply. *Nature*, **482**, E5-E6.
He, F. & Hubbell, S. (2013) Estimating extinction from species–area relationships: why the numbers do not add up. *Ecology*, **94**, 1905-1912.
He, F., Gaston, K.J. & Wu, J. (2002) On species occupancy-abundance models. *Ecoscience*, **9**, 119-126.
Heaney, L.R. (2000) Dynamic disequilibrium: a long-term, large-scale perspective on the equilibrium model of island biogeography. *Global Ecology and Biogeography*, **9**, 59-74.
Hedges, L.V. & Olkin, I. (1985) *Statistical methods for meta-analysis*. Academic Press, London.
Heino, J. (2005) Positive relationship between regional distribution and local abundance in stream insects: a consequence of niche breadth or niche position? *Ecography*, **28**, 345-354.
Helzer, C.J. & Jelinski, D.E. (1999) The relative importance of patch area and perimeter–area ratio to grassland breeding birds. *Ecological Applications*, **9**, 1448-1458.
Hengeveld, R. & Haeck, J. (1982) The distribution of abundance. I. Measurements. *Journal of Biogeography*, **9**, 303-316.
Henle, K., Davies, K.F., Kleyer, M., Margules, C. & Settele, J. (2004) Predictors of species sensitivity to fragmentation. *Biodiversity and Conservation*, **13**, 207-251.
Hill, J.K., Hamer, K.C., Lace, K.C. & Banham, W.M.T. (1995) Effects of selective logging on tropical forest butterflies on Buru, Indonesia. *Journal of Applied Ecology*, **32**, 754-760.
Hill, J.K., Thomas, C.D. & Huntley, B. (1999) Climate and habitat availability determine 20th century changes in a butterfly's range margin. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **266**, 1197-1206.

Hill, J.K., Collingham, Y.C., Thomas, C.D., Blakeley, D.S., Fox, R., Moss, D. & Huntley, B. (2001) Impacts of landscape structure on butterfly range expansion. *Ecology Letters*, **4**, 313-321.

Hinsley, S.A., Bellamy, P.E., Newton, I. & Sparks, T.H. (1994) Factors influencing the presence of individual breeding bird species in woodland fragments. In: *English Nature Research Reports 99*.

Hinsley, S.A., Bellamy, P.E., Newton, I. & Sparks, T.H. (1995) Habitat and landscape factors influencing the presence of individual breeding bird species in woodland fragments. *Journal of Avian Biology*, **26**, 94-104.

Hinsley, S.A., Bellamy, P.E., Newton, I. & Sparks, T.H. (1996) Influences of population size and woodland area on bird species distributions in small woods. *Oecologia*, **105**, 100-106.

Holling, C.S. (1992) Cross-scale morphology, geometry, and dynamics of ecosystems. *Ecological Monographs*, **62**, 447-502.

Holt, R.D., Lawton, J.H., Gaston, K.J. & Blackburn, T.M. (1997) On the relationship between range size and local abundance: back to basics. *Oikos*, **78**, 183-190.

Holt, R.D., Lawton, J.H., Polis, G.A. & Martinez, N.D. (1999) Trophic rank and the species–area relationship. *Ecology*, **80**, 1495-1504.

Honnay, O., Hermy, M. & Coppin, P. (1999) Nested plant communities in deciduous forest fragments: species relaxation or nested habitats? *Oikos*, **84**, 119-129.

Horner-Devine, M.C., Lage, M., Hughes, J.B. & Bohannan, B.J.M. (2004) A taxa-area relationship for bacteria. *Nature*, **432**, 750-753.

Houlahan, J.E., Currie, D.J., Cottenie, K., Cumming, G.S., Ernest, S.K.M., Findlay, C.S., Fuhlendorf, S.D., Gaedke, U., Legendre, P., Magnuson, J.J., McArdle, B.H., Muldavin, E.H., Noble, D., Russell, R., Stevens, R.D., Willis, T.J., Woiwod, I.P. & Wondzell, S.M. (2007) Compensatory dynamics are rare in natural ecological communities. *Proceedings of the National Academy of Sciences USA*, **104**, 3273-3277.

Howe, R.W. (1984) Local dynamics of bird assemblages in small forest habitat islands in Australia and North America. *Ecology*, **65**, 1585-1601.

Hu, G., Wu, J., Feeley, K.J., Xu, G. & Yu, M. (2012) The effects of landscape variables on the species-area relationship during late-stage habitat fragmentation. *PLOS ONE*, **7**, e43894.

Hubbell, S.P. (2001) *The unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton.

Hubbell, S.P., Condit, R. & Foster, R.B. (2005) *Barro Colorado Forest census plot data.* Available at: https://ctfs.arnarb.harvard.edu/webatlas/datasets/bci. (accessed April 2013).

Huggett, A.J. (2005) The concept and utility of 'ecological thresholds' in biodiversity conservation. *Biological Conservation*, **124**, 301-310.

Hugueny, B., Movellan, A. & Belliard, J. (2011) Habitat fragmentation and extinction rates within freshwater fish communities: a faunal relaxation approach. *Global Ecology and Biogeography*, **20**, 449-463.

Hui, C. (2008) On species–area and species accumulation curves: A comment on Chong and Stohlgren's index. *Ecological Indicators*, **8**, 327-329.

Humphreys, W.F. & Kitchener, D.J. (1982) The effect of habitat utilisation on species-area curves: implications for optimal reserve area. *Journal of Biogeography*, **9**, 381-396.

Hutchinson, G.E. (1957) Concluding remarks. *Cold Spring Harb. Symp. Quant. Biol.*, **22**, 415-427.

Jamoneau, A., Sonnier, G., Chabrerie, O., Closset-Kopp, D., Saguez, R., Gallet-Moron, E. & Decocq, G. (2011) Drivers of plant species assemblages in forest patches among contrasted dynamic agricultural landscapes. *Journal of Ecology*, **99**, 1152-1161.

Janzen, D.H. (1986) The eternal external threat. *Conservation biology* (ed. by M.E. Soulé), pp. 286-303. Sinauer Associates, Sunderland, Massachusetts.

Jaynes, E.T. (2003) *Probability theory: the logic of science*. Cambridge University Press, New York.

Johnstone, C.P., Lill, A. & Reina, R.D. (2014) Habitat loss, fragmentation and degradation effects on small mammals: Analysis with conditional inference tree statistical modelling. *Biological Conservation*, **176**, 80-98.

Joly, P., Morand, C. & Cohas, A. (2003) Habitat fragmentation and amphibian conservation: building a tool for assessing landscape matrix connectivity. *Comptes Rendus Biologies*, **326**, S132-S139.

Jonsson, M., Englund, G. & Wardle, D.A. (2011) Direct and indirect effects of area, energy and habitat heterogeneity on breeding bird communities. *Journal of Biogeography*, **38**, 1186-1196.

Jules, E.S. & Shahani, P. (2003) A broader ecological context to habitat fragmentation: Why matrix habitat is more important than we thought. *Journal of Vegetation Science*, **14**, 459-464.

Kadmon, R. (1995) Nested species subsets and geographic isolation: a case study. *Ecology*, **76**, 458-465.

Kalmar, A. & Currie, D.J. (2006) A global model of island biogeography. *Global Ecology and Biogeography*, **15**, 72-81.

Karanth, K.K., Nichols, J.D., Sauer, J.R. & Hines, J.E. (2006) Comparative dynamics of avian communities across edges and interiors of North American ecoregions. *Journal of Biogeography*, **33**, 674-682.

Karr, J.R. (1982) Population variability and extinction in the avifauna of a tropical land bridge island. *Ecology*, **63**, 1975-1978.

Kim, D.-H., Cho, W.-S. & Chon, T.-S. (2013) Self-organizing map and species abundance distribution of stream benthic macroinvertebrates in revealing community patterns in different seasons. *Ecological Informatics*, **17**, 14-29.

Kipp, F. (1959) Der Handflugel-Index als flugbiologische Mab. *Vogelwarte*, **20**, 77–86.

Kisel, Y., McInnes, L., Toomey, N.H. & Orme, C.D.L. (2011) How diversification rates and diversity limits combine to create large-scale species–area relationships. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **366**, 2514-2525.

Kitchener, D.J., Chapman, A., Muir, B.G. & Palmer, M. (1980) The conservation value for mammals of reserves in the western Australian wheatbelt. *Biological Conservation*, **18**, 179-207.

Kitzes, J. & Harte, J. (2014) Beyond the species–area relationship: improving macroecological extinction estimates. *Methods in Ecology and Evolution*, **5**, 1-8.

Koh, L.P. & Ghazoul, J. (2010) A matrix-calibrated species-area model for predicting biodiversity losses due to land-use change. *Conservation Biology*, **24**, 994-1001.

Koh, L.P., Sodhi, N.S. & Brook, B.W. (2004) Ecological correlates of extinction proneness in tropical butterflies. *Conservation Biology*, **18**, 1571-1578.

Koh, L.P., Lee, T.M., Sodhi, N.S. & Ghazoul, J. (2010) An overhaul of the species–area approach for predicting biodiversity loss: incorporating matrix and edge effects. *Journal of Applied Ecology*, **47**, 1063-1070.

Krauss, J., Bommarco, R., Guardiola, M., Heikkinen, R.K., Helm, A., Kuussaari, M., Lindborg, R., Öckinger, E., Pärtel, M., Pino, J., Pöyry, J., Raatikainen, K.M., Sang,

A., Stefanescu, C., Teder, T., Zobel, M. & Steffan-Dewenter, I. (2010) Habitat fragmentation causes immediate and time-delayed biodiversity loss at different trophic levels. *Ecology Letters*, **13**, 597-605.
Kruess, A. & Tscharntke, T. (1994) Habitat fragmentation, species loss, and biological control. *Science*, **264**, 1581-1584.
Labra, F.A., Abades, S.R. & Marquet, P.A. (2005) Distribution and abundance: scaling patterns in exotic and native bird species. *Species invasions: insights into ecology, evolution, and biogeography* (ed. by D.F. Sax, J.J. Stachowicz and S.D. Gaines), pp. 421-446. Sinauer, Sunderland, MA.
Lack, D. (1976) *Island biology illustrated by the land birds of Jamaica*. University of California Press, Berkeley.
Ladle, R.J. & Whittaker, R.J. (eds.) (2011) *Conservation biogeography*. Wiley-Blackwell, Chichester.
Laliberté, E. & Shipley, B. (2011) *FD: measuring functional diversity from multiple traits, and other tools for functional ecology*.
Lambeck, R.J. (1997) Focal species: a multi-species umbrella for nature conservation. *Conservation Biology*, **11**, 849-856.
Lambshead, P.J.D. (1986) Sub-catastrophic sewage and industrial waste contamination as revealed by marine nematode faunal analysis. *Marine Ecology Progress Series*, **29**, 247–260.
Lambshead, P.J.D. & Boucher, G. (2003) Marine nematode deep-sea biodiversity – hyperdiverse or hype? *Journal of Biogeography*, **30**, 475-485.
Lambshead, P.J.D., Platt, H.M. & Shaw, K.M. (1983) The detection of differences among assemblages of marine benthic species based on an assessment of dominance and diversity. *Journal of Natural History*, **17**, 859-874.
Laurance, W.F. (2008) Theory meets reality: how habitat fragmentation research has transcended island biogeographic theory. *Biological Conservation*, **141**, 1731-1744.
Laurance, W.F. & Yensen, E. (1991) Predicting the impacts of edge effects in fragmented habitats. *Biological Conservation*, **55**, 77-92.
Laurance, W.F. & Bierregaard, J., Richard O. (1996) Fragmented tropical forests. *Bulletin of the Ecological Society of America*, **77**, 34-36.
Laurance, W.F. & Bierregaard, R.O. (eds.) (1997) *Tropical forest remnants: ecology, management, and conservation of fragmented communities*. University of Chicago Press, Chicago.
Laurance, W.F. & Peres, C.A. (eds.) (2006) *Emerging threats to tropical forests*. University of Chicago Press, Chicago.
Laurance, W.F., Ferreira, L.V., Rankin-de Merona, J.M. & Laurance, S.G. (1998) Rainforest fragmentation and the dynamics of Amazonian tree communities. *Ecology*, **79**, 2032-2040.
Laurance, W.F., Lovejoy, T.E., Vasconcelos, H.L., Bruna, E.M., Didham, R.K., Stouffer, P.C., Gascon, C., Bierregaard, R.O., Laurance, S.G. & Sampaio, E. (2002) Ecosystem decay of Amazonian forest fragments: a 22-year investigation. *Conservation Biology*, **16**, 605-618.
Laurance, W.F., Goosem, M. & Laurance, S.G.W. (2009) Impacts of roads and linear clearings on tropical forests. *Trends in Ecology & Evolution*, **24**, 659-669.
Lawton, J.H. (1999) Are there general laws in ecology? *Oikos*, **84**, 177-192.
Lees, A.C. & Peres, C.A. (2008) Avian life-history determinants of local extinction risk in a hyper-fragmented neotropical forest landscape. *Animal Conservation*, **11**, 128-137.
Lees, A.C. & Peres, C.A. (2009) Gap-crossing movements predict species occupancy in Amazonian forest fragments. *Oikos*, **118**, 280-290.

Leibold, M.A. & Mikkelson, G.M. (2002) Coherence, species turnover, and boundary clumping: elements of meta-community structure. *Oikos*, **97**, 237-250.
Levenson, J.B. (1981) Woodplots as biogeographic islands in southern Wisconsin. Forest island dynamics in man-dominated landscapes (ed. by R.L. Burgess and D.M. Sharpe), pp. 13-39. Springer-Verlag, New-York.
Levins, R. (1969) Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the ESA*, **15**, 237-240.
Lindenmayer, D.B. & Luck, G. (2005) Synthesis: thresholds in conservation and management. *Biological Conservation*, **124**, 351-354.
Lindenmayer, D.B. & Fischer, J. (2006) *Habitat fragmentation and landscape change: an ecological and conservation synthesis*. Island Press, Washington DC.
Lindenmayer, D.B. & Fischer, J. (2007) Tackling the habitat fragmentation panchreston. *Trends in Ecology and Evolution*, **22**, 127-132.
Lindenmayer, D.B., Fischer, J. & Cunningham, R.B. (2005) Native vegetation cover thresholds associated with species responses. *Biological Conservation*, **124**, 311-316.
Liu, J.-J. & Slik, J.W.F. (2014) Forest fragment spatial distribution matters for tropical tree conservation. *Biological Conservation*, **171**, 99-106.
Locey, K.J. & White, E.P. (2013) How species richness and total abundance constrain the distribution of abundance. *Ecology Letters*, **16**, 1177-1185.
Lomolino, M.V. (1982) Species-area and species-distance relationships of terrestrial mammals in the Thousand Island Region. *Oecologia*, **54**, 72-75.
Lomolino, M.V. (2000) Ecology's most general, yet protean pattern: the species-area relationship. *Journal of Biogeography*, **27**, 17-26.
Lomolino, M.V. (2002) The species-area relationship does not have an asymptote - Comment. *Journal of Biogeography*, **29**, 555-557.
Lomolino, M.V. & Weiser, M.D. (2001) Towards a more general species-area relationship: diversity on all islands, great and small. *Journal of Biogeography*, **28**, 431-445.
Lomolino, M.V., Brown, J.H. & Davis, R. (1989) Island biogeography of montane forest mammals in the American Southwest. *Ecology*, **70**, 180-194.
Lomolino, M.V., Riddle, B.R., Whittaker, R.J. & Brown, J.H. (2010a) *Biogeography*, Fourth edn. Sunderland MA, Sinauer Associates.
Lomolino, M.V., Brown, J.H. & Sax, D.F. (2010b) Island biogeography theory: reticulations and reintegration of a "biogeography of the species". *The theory of island biogeography revisited* (ed. by J.B. Losos and R.E. Ricklefs). Princeton University Press, Princeton.
Losos, J.B. & Schluter, D. (2000) Analysis of an evolutionary species-area relationship. *Nature*, **408**, 847-850.
Lövei, G.L., Magura, T., Tothmeresz, B. & Kodobocz, V. (2006) The influence of matrix and edges on species richness patterns of ground beetles (Coleoptera : Carabidae) in habitat islands. *Global Ecology and Biogeography*, **15**, 283-289.
Lovejoy, T.E., Bierregaard Jr., R.O., Rylands, A.B., Malcolm, J.R., Quintela, C.E., Harper, L.H., Brown Jr., K.S., Powell, A.H., Powell, G.V.N., Schubart, H.O.R. & Hays, M. (1986) Edge and other effects of isolation on Amazon forest fragments. *Conservation biology: the science of scarcity and diversity*. (ed. by M.E. Soulé), pp. 185-257. Sinauer Associates, Sunderland.
MacArthur, R.H. (1957) On the relative abundance of bird species. *Proceedings of the National Academy of Sciences USA*, **43**, 293-295.
MacArthur, R.H. (1960) On the relative abundance of species. *The American Naturalist*, **94**, 25-36.
MacArthur, R.H. (1972) *Geographical ecology*. Princeton University Press, Princeton.

MacArthur, R.H. & MacArthur, J.W. (1961) On bird species diversity. *Ecology*, **42**, 594-598.
MacArthur, R.H. & Wilson, E.O. (1963) An equilibrium theory of insular zoogeography. *Evolution*, **17**, 373-387.
MacArthur, R.H. & Wilson, E.O. (1967) *The theory of island biogeography*. Princeton University Press, Princeton.
MacArthur, R.H., Diamond, J.M. & Karr, J.R. (1972) Density compensation in island faunas. *Ecology*, **53**, 330-342.
MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Andrew Royle, J. & Langtimm, C.A. (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, **83**, 2248-2255.
MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L. & Hines, J.E. (2006) *Occupancy estimation and modelling*. Academic Press, San Diego, CA.
MacNally, R. (2007) Use of the abundance spectrum and relative abundance distributions to analyze assemblage change in massively altered landscapes. *The American Naturalist*, **170**, 319-330.
Magura, T., Ködöböcz, V. & Tóthmérész, B. (2001) Effects of habitat fragmentation on carabids in forest patches. *Journal of Biogeography*, **28**, 129-138.
Magura, T., Báldi, A. & Horváth, R. (2008) Breakdown of the species-area relationship in exotic but not in native forest patches. *Acta Oecologica*, **33**, 272-279.
Magurran, A.E. (2004) *Measuring biological diversity*. Blackwell, Oxford.
Magurran, A.E. (2007) Species abundance distributions over time. *Ecology Letters*, **10**, 347-354.
Magurran, A.E. & Henderson, P.A. (2003) Explaining the excess of rare species in natural species abundance distributions. *Nature*, **422**, 714-716.
Malanson, G.P. (2008) Extinction debt: origins, developments, and applications of a biogeographical trope. *Progress in Physical Geography*, **32**, 277-292.
Manning, A.D., Lindenmayer, D.B. & Nix, H.A. (2004) Continua and Umwelt: novel perspectives on viewing landscapes. *Oikos*, **104**, 621-628.
Margules, C.R. & Pressey, R.L. (2000) Systematic conservation planning. *Nature*, **405**, 243-253.
Marini, M.Â. (2001) Effects of forest fragmentation on birds of the cerrado region, Brazil. *Bird Conservation International*, **11**, 13-25.
Maron, M., Bowen, M., Fuller, R.A., Smith, G.C., Eyre, T.J., Mathieson, M., Watson, J.E.M. & McAlpine, C.A. (2012) Spurious thresholds in the relationship between species richness and vegetation cover. *Global Ecology and Biogeography*, **21**, 682-692.
Marquet, P.A., Keymer, J.A. & Cofre, H. (2003) Breaking the stick in space: of niche models, metacommunities and patterns in the relative abundance of species. *Macroecology: concepts and consequences* (ed. by T.M. Blackburn and K.J. Gaston), pp. 64-86. Blackwell Oxford.
Marquet, P.A., Fernández, M., Navarrete, S.A. & Valdovinos, C. (2004) Diversity emerging: toward a deconstruction of biodiversity patterns. *Frontiers of biogeography: new directions in the geography of nature* (ed. by M.V. Lomolino and L.R. Heaney), pp. 191-209. Sinauer, Sunderland, MA.
Mason, C. (2001) Woodland area, species turnover and the conservation of bird assemblages in lowland England. *Biodiversity and Conservation*, **10**, 495-510.
Matias, M.G., Gravel, D., Guilhaumon, F., Desjardins-Proulx, P., Loreau, M., Münkemüller, T. & Mouquet, N. (2014) Estimates of species extinctions from species–area relationships strongly depend on ecological context. *Ecography*, **37**, 431-442.

Matthews, T.J. & Whittaker, R.J. (2014a) On the species abundance distribution in applied ecology and biodiversity management. *Journal of Applied Ecology*, doi: 10.1111/1365-2664.12380

Matthews, T.J. & Whittaker, R.J. (2014b) Fitting and comparing competing models of the species abundance distribution: assessment and prospect. *Frontiers of Biogeography*, **6**, 67-82.

Matthews, T.J. & Whittaker, R.J. (2014c) Neutral theory and the species abundance distribution: recent developments and prospects for unifying niche and neutral perspectives. *Ecology and Evolution*, **4**, 2263–2277.

Matthews, T.J., Cottee-Jones, H.E. & Whittaker, R.J. (2014a) Habitat fragmentation and the species–area relationship: a focus on total species richness obscures the impact of habitat loss on habitat specialists. *Diversity and Distributions*, **20**, 1136-1146.

Matthews, T.J., Steinbauer, M.J., Tzirkalli, E., Triantis, K.A. & Whittaker, R.J. (2014b) Thresholds and the species–area relationship: a synthetic analysis of habitat island datasets. *Journal of Biogeography*, **41**, 1018-1028.

Matthews, T.J., Borges, P.A.V. & Whittaker, R.J. (2014c) Multimodal species abundance distributions: a deconstruction approach reveals the processes behind the pattern. *Oikos*, **123**, 533-544.

May, R.M. (1975) Patterns of species abundance and diversity. *Ecology and evolution of communities* (ed. by M.L. Cody and J.M. Diamond), pp. 81-120. Harvard University Press, Cambridge, MA.

Mazel, F., Guilhaumon, F., Mouquet, N., Devictor, V., Gravel, D., Renaud, J., Cianciaruso, M.V., Loyola, R., Diniz-Filho, J.A.F., Mouillot, D. & Thuiller, W. (2014) Multifaceted diversity–area relationships reveal global hotspots of mammalian species, trait and lineage diversity. *Global Ecology and Biogeography*, **23**, 836-847.

Mazerolle, M. (2011) *AICcmodavg: model selection and multimodel inference based on (Q)AIC(c)*. R package version 1.20.

McAbendroth, L., Foggo, A., Rundle, S.D. & Bilton, D.T. (2005) Unravelling nestedness and spatial pattern in pond assemblages. *Journal of Animal Ecology*, **74**, 41-49.

McCollin, D. (1993) Avian distribution patterns in a fragmented wooded landscape (North Humberside, U.K.): the role of between-patch and within-patch structure. *Global Ecology and Biogeography Letters*, **3**, 48-62.

McCoy, E.D. & Mushinsky, H.R. (2007) Estimates of minimum patch size depend on the method of estimation and the condition of the habitat. *Ecology*, **88**, 1401-1407.

McGill, B. (2003) Strong and weak tests of macroecological theory. *Oikos*, **102**, 679-685.

McGill, B.J. (2011) Species abundance distributions. *Biological diversity: frontiers in measurement and assessment* (ed. by A.E. Magurran and B.J. Mcgill), pp. 105-122. Oxford University Press, Oxford.

McGill, B.J., Enquist, B.J., Weiher, E. & Westoby, M. (2006) Rebuilding community ecology from functional traits. *Trends in Ecology & Evolution*, **21**, 178-185.

McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K., Dornelas, M., Enquist, B.J., Green, J.L., He, F., Hurlbert, A.H., Magurran, A.E., Marquet, P.A., Maurer, B.A., Ostling, A., Soykan, C.U., Ugland, K.I. & White, E.P. (2007) Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters*, **10**, 995-1015.

McIntyre, S. & Hobbs, R. (1999) A framework for conceptualizing human effects on landscapes and its relevance to management and research models. *Conservation Biology*, **13**, 1282-1292.

Meijer, S., Whittaker, R.J. & Borges, P.A.V. (2011) The effects of land-use change on arthropod richness and abundance on Santa Maria Island (Azores): unmanaged plantations favour endemic beetles. *Journal of Insect Conservation*, **15**, 505-522.

Melo, A.S. (2012) *Community ecology analyses*. R package version 1.5.9.

Melo, A.S., Cianciaruso, M.V. & Almeida-Neto, M. (2014) treeNODF: nestedness to phylogenetic, functional and other tree-based diversity metrics. *Methods in Ecology and Evolution*, **5**, 563-572.

Mendenhall, C.D., Karp, D.S., Meyer, C.F.J., Hadly, E.A. & Daily, G.C. (2014a) Predicting biodiversity change and averting collapse in agricultural landscapes. *Nature*, **509**, 213-217.

Mendenhall, C.D., Frishkoff, L.O., Santos-Barrera, G., Pacheco, J., Mesfun, E., Quijano, F.M., Ehrlich, P.R., Ceballos, G., Daily, G.C. & Pringle, R.M. (2014b) Countryside biogeography of Neotropical reptiles and amphibians. *Ecology*, **95**, 856-870.

Met Office (2014) *Guildford climate*. Available at: http://www.metoffice.gov.uk/public/weather/climate/gcpe7p302 (accessed August 2014).

Miklós, I. & Podani, J. (2004) Randomization of presence–absence matrices: comments and new algorithms. *Ecology*, **85**, 86-92.

Miles, D.B., Ricklefs, R.E. & Travis, J. (1987) Concordance of ecomorphological relationships in three assemblages of passerine birds. *The American Naturalist*, **129**, 347-364.

Millar, R.B., Anderson, M.J. & Tolimieri, N. (2011) Much ado about nothings: using zero similarity points in distance-decay curves. *Ecology*, **92**, 1717-1722.

Moilanen, A. (2002) Implications of empirical data quality to metapopulation model parameter estimation and application. *Oikos*, **96**, 516-530.

Monnet, A.-C., Jiguet, F., Meynard, C.N., Mouillot, D., Mouquet, N., Thuiller, W. & Devictor, V. (2014) Asynchrony of taxonomic, functional and phylogenetic diversity in birds. *Global Ecology and Biogeography*, **23**, 780-788.

Montoya, D., Alburquerque, F.S., Rueda, M. & Rodriguez, M.A. (2010) Species' response patterns to habitat fragmentation: do trees support the extinction threshold hypothesis? *Oikos*, **119**, 1335-1343.

Moore, N.W. & Hooper, M.D. (1975) On the number of bird species in British woods. *Biological Conservation*, **8**, 239-250.

Morrison, L.W. (2013) Nestedness in insular floras: spatiotemporal variation and underlying mechanisms. *Journal of Plant Ecology*, **6**, 480-488.

Mortelliti, A., Fagiani, S., Battisti, C., Capizzi, D. & Boitani, L. (2010) Independent effects of habitat loss, habitat fragmentation and structural connectivity on forest-dependent birds. *Diversity and Distributions*, **16**, 941-951.

Motomura, I. (1932) On the statistical treatment of communities *Zoological Magazine Tokyo*, **44**, 379-383.

Mouchet, M., Guilhaumon, F., Villéger, S., Mason, N.W.H., Tomasini, J.-A. & Mouillot, D. (2008) Towards a consensus for calculating dendrogram-based functional diversity indices. *Oikos*, **117**, 794-800.

Mouillot, D. & Lepretre, A. (2000) Introduction of relative abundance distribution (RAD) indices, estimated from the rank-frequency diagrams (RFD), to assess changes in community diversity. *Environmental Monitoring and Assessment*, **63**, 279-295.

Mouillot, D. & Wilson, J.B. (2002) Can we tell how a community was constructed? A comparison of five evenness indices for their ability to identify theoretical models of community construction. *Theoretical Population Biology*, **61**, 141-151.

Mouillot, D., George-Nascimento, M. & Poulin, R. (2003) How parasites divide resources: a test of the niche apportionment hypothesis. *Journal of Animal Ecology*, **72**, 757-764.
Mouillot, D., Graham, N.A.J., Villéger, S., Mason, N.W.H. & Bellwood, D.R. (2013) A functional approach reveals community responses to disturbances. *Trends in Ecology & Evolution*, **28**, 167-177.
Myers, N. (1986) Tropical deforestation and a mega-extinction spasm. *Conservation biology* (ed. by M.E. Soulé), pp. 394-409. Sinauer Associates, Sunderland, Massachusetts.
Naeem, S., Duffy, J.E. & Zavaleta, E. (2012) The functions of biological diversity in an age of extinction. *Science*, **336**, 1401-1406.
Naveh, Z. & Lieberman, A.S. (1984) *Landscape ecology*. Springer-Verlag, New York.
Newbold, T., Scharlemann, J.P.W., Butchart, S.H.M., Şekercioğlu, Ç.H., Alkemade, R., Booth, H. & Purves, D.W. (2013) Ecological traits affect the response of tropical forest bird species to land-use intensity. *Proceedings of the Royal Society B: Biological Sciences*, **280**
Newbold, T., Hudson, L.N., Phillips, H.R.P., Hill, S.L.L., Contu, S., Lysenko, I., Blandon, A., Butchart, S.H.M., Booth, H.L., Day, J., De Palma, A., Harrison, M.L.K., Kirkpatrick, L., Pynegar, E., Robinson, A., Simpson, J., Mace, G.M., Scharlemann, J.P.W. & Purvis, A. (2014) A global model of the response of tropical and sub-tropical forest biodiversity to anthropogenic pressures. *Proceedings of the Royal Society B: Biological Sciences*, **281**
Newmark, W.D. (1986) Species are relationship and its determinants for mammals in western north-american national-parks. *Biological Journal of the Linnean Society*, **28**, 83-98.
Newsome, A.E. & Catling, P.C. (1979) Habitat preferences of mammals inhabiting heathlands of warm, temperate montane and coastal and alpine regions of southeastern Australia. *Heathlands and related shrublands of the world* (ed. by R.L. Specht), pp. 301–316. Elsevier, Amsterdam.
Nicholson, E., Lindenmayer, D.B., Frank, K. & Possingham, H.P. (2013) Testing the focal species approach to making conservation decisions for species persistence. *Diversity and Distributions*, **19**, 530-540.
Numata, M. (1950) The homogeneity of plant communities– studies on the structure of plant communities VI (in Japanese). *Shokubutsugaku Zasshi*, **63**, 747–748.
Nummelin, M. (1998) Log-normal distribution of species abundances is not a universal indicator of rain forest disturbance. *Journal of Applied Ecology*, **35**, 454-457.
Oksanen, J. (2004) *Incidence function model in R*. Available at: http://cc.oulu.fi/~jarioksa/opetus/openmeta/metafit.pdf (accessed February 2001).
Oksanen, J.F., Blanchet, G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B. & Simpson, G.L. (2013) *"Package 'vegan'.* R package version 2.0-9.
Opdam, P. (1991) Metapopulation theory and habitat fragmentation: a review of holarctic breeding bird studies. *Landscape Ecology*, **5**, 93-106.
Ostfeld, R.S. & LoGiudice, K. (2003) Community disassembly, biodiversity loss, and the erosion of an ecosystem service. *Ecology*, **84**, 1421-1427.
Ovaskainen, O. (2002) Long-term persistence of species and the SLOSS problem. *Journal of Theoretical Biology*, **218**, 419-433.
Paradis, E., Baillie, S.R., Sutherland, W.J. & Gregory, R.D. (1998) Patterns of natal and breeding dispersal in birds. *Journal of Animal Ecology*, **67**, 518-536.
Paradis, E., Claude, J. & Strimmer, K. (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289-290.
Pardini, R., Bueno, A.d.A., Gardner, T.A., Prado, P.I. & Metzger, J.P. (2010) Beyond the fragmentation threshold hypothesis: regime shifts in biodiversity across fragmented landscapes. *PLOS ONE*, **5**, e13666.

Patterson, B.D. & Atmar, W. (1986) Nested subsets and the structure of insular mammalian faunas and archipelagos. *Biological Journal of the Linnean Society*, **28**, 65-82.
Patton, D.R. (1975) A diversity index for quantifying habitat "edge". *Wildlife Society Bulletin*, **3**, 171-173.
Pavlacky, D.C. & Anderson, S.H. (2007) Does avian species richness in natural patch mosaics follow the forest fragmentation paradigm? *Animal Conservation*, **10**, 57-68.
Pe'er, G., Tsianou, M.A., Franz, K.W., Matsinos, Y.G., Mazaris, A.D., Storch, D., Kopsova, L., Verboom, J., Baguette, M., Stevens, V.M. & Henle, K. (2014) Toward better application of minimum area requirements in conservation planning. *Biological Conservation*, **170**, 92-102.
Pearce, J. & Ferrier, S. (2001) The practical value of modelling relative abundance of species for regional conservation planning: a case study. *Biological Conservation*, **98**, 33-43.
Pereira, H.M. & Daily, G.C. (2006) Modelling biodiversity dynamics in countryside landscapes. *Ecology*, **87**, 1877-1885.
Pereira, H.M., Borda-de-Água, L. & Martins, I.S. (2012) Geometry and scale in species-area relationships. *Nature*, **482**, E3-E4.
Pereira, H.M., Ziv, G.U.Y. & Miranda, M. (2014) Countryside species–area relationship as a valid alternative to the matrix-calibrated species–area model. *Conservation Biology*, **28**, 874-876.
Petchey, O.L. & Gaston, K.J. (2002) Functional diversity (FD), species richness and community composition. *Ecology Letters*, **5**, 402-411.
Petchey, O.L. & Gaston, K.J. (2006) Functional diversity: back to basics and looking forward. *Ecology Letters*, **9**, 741-758.
Petchey, O.L., Hector, A. & Gaston, K.J. (2004) How do different measures of functional diversity perform? *Ecology*, **85**, 847-857.
Pickett, S.T.A., Parker, V.T. & Fiedler, P.L. (1992) The new paradigm in ecology: implications for conservation biology above the species level. *Conservation biology; the theory and practice of nature preservation and management* (ed. by P.L. Fiedler and S.K. Jain), pp. 65-88 Chapman and Hall, New York.
Pielou, E.C. (1969) *An introduction to mathematical ecology*. John Wiley & Sons, New York.
Pielou, E.C. (1975) *Ecological diversity*. Wiley, New York.
Poschlod, P., Kiefer, S., Tränkle, U., Fischer, S. & Bonn, S. (1998) Plant species richness in calcareous grasslands as affected by dispersability in space and time. *Applied Vegetation Science*, **1**, 75-91.
Poulin, R. & Guégan, J.-F. (2000) Nestedness, anti-nestedness, and the relationship between prevalence and intensity in ectoparasite assemblages of marine fish: a spatial model of species coexistence. *International Journal for Parasitology*, **30**, 1147-1152.
Preston, F.W. (1948) The commonness, and rarity, of species. *Ecology*, **29**, 254-283.
Preston, F.W. (1962) Canonical distribution of commoness and rarity.1. *Ecology*, **43**, 185-215.
Price, J.P. & Clague, D.A. (2002) How old is the Hawaiian biota? Geology and phylogeny suggest recent divergence. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **269**, 2429-2435.
Primack, R.B. (2010) *Essentials of conservation biology*, Fifth edn. Macmillan Science, Basingstoke.
Prugh, L.R. (2009) An evaluation of patch connectivity measures. *Ecological Applications*, **19**, 1300-1310.

Prugh, L.R., Hodges, K.E., Sinclair, A.R.E. & Brashares, J.S. (2008) Effect of habitat area and isolation on fragmented animal populations. *Proceedings of the National Academy of Sciences USA*, **105**, 20770-20775.

Pueyo, S., He, F. & Zillio, T. (2007) The maximum entropy formalism and the idiosyncratic theory of biodiversity. *Ecology Letters*, **10**, 1017-1028.

Pulliam, H.R. (1988) Sources, sinks, and population regulation. *The American Naturalist*, **132**, 652-661.

Qie, L., Lee, T.M., Sodhi, N.S. & Lim, S.L.H. (2011) Dung beetle assemblages on tropical land-bridge islands: small island effect and vulnerable species. *Journal of Biogeography*, **38**, 792-804.

Quinn, J. & Harrison, S. (1988) Effects of habitat fragmentation and isolation on species richness: evidence from biogeographic patterns. *Oecologia*, **75**, 132-140.

R Development Core Team (2013) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria.

Rabinowitz, D., Cairns, S. & Dillon, T. (1986) Seven forms of rarity and their frequency in the flora in the British Isles. *Conservation biology: the science of scarcity and diversity* (ed. by M.E. Soulé), pp. 182-204. Sinauer Associates, Sunderland, Mass.

Radford, J.Q., Bennett, A.F. & Cheers, G.J. (2005) Landscape-level thresholds of habitat cover for woodland-dependent birds. *Biological Conservation*, **124**, 317-337.

Rapoport, E.H. (1982) *Areography: geographical strategies of species*. Pergamon Press, Oxford.

Reed, D.H., O'Grady, J.J., Brook, B.W., Ballou, J.D. & Frankham, R. (2003) Estimates of minimum viable population sizes for vertebrates and factors influencing those estimates. *Biological Conservation*, **113**, 23-34.

Ribeiro, S.P., Borges, P.A.V., Gaspar, C., Melo, C., Serrano, A.R.M., Amaral, J., Aguiar, C., André, G. & Quartau, J.A. (2005) Canopy insect herbivores in the Azorean Laurisilva forests: key host plant species in a highly generalist insect community. *Ecography*, **28**, 315-330.

Ricketts, T.H. (2001) The matrix matters: effective isolation in fragmented landscapes. *The American Naturalist*, **158**, 87-99.

Ricklefs, R.E. & Lovette, I.J. (1999) The roles of island area per se and habitat diversity in the species–area relationships of four Lesser Antillean faunal groups. *Journal of Animal Ecology*, **68**, 1142-1160.

Ricklefs, R.E. & Bermingham, E. (2004) History and the species-area relationship in Lesser Antillean birds. *The American Naturalist*, **163**, 227-239.

Ries, L., Jr, R.J.F., Battin, J. & Sisk, T.D. (2004) Ecological responses to habitat edges: mechanisms, models, and variability explained. *Annual Review of Ecology, Evolution, and Systematics*, **35**, 491-522.

Rigal, F., Whittaker, R.J., Triantis, K.A. & Borges, P.A.V. (2013) Integration of non-indigenous species within the interspecific abundance-occupancy relationship. *Acta Oecologica*, **48**, 69-75.

Risk, B.B., de Valpine, P. & Beissinger, S.R. (2011) A robust-design formulation of the incidence function model of metapopulation dynamics applied to two species of rails. *Ecology*, **92**, 462-474.

Robinson, W.D. (1999) Long-term changes in the avifauna of Barro Colorado Island, Panama, a tropical forest isolate. *Conservation Biology*, **13**, 85-97.

Robinson, W.D. & Sherry, T. (2012) Mechanisms of avian population decline and species loss in tropical forest fragments. *Journal of Ornithology*, **153**, 141-152.

Rosa, I.M.D., Souza, C. & Ewers, R.M. (2012) Changes in size of deforested patches in the Brazilian Amazon. *Conservation Biology*, **26**, 932-937.

Rosenzweig, M.L. (1995) *Species diversity in space and time*. Cambridge University Press, Cambridge.

Rosenzweig, M.L., Donoghue, J., Li, Y.M. & Yuan, C. (2011) Estimating species density. *Biological diversity: frontiers in measurement and assessment* (ed. by A.E. Magurran and B.J. Mcgill), pp. 276-288. Oxford University Press, Oxford.

Rosindell, J. & Harmon, L.J. (2013) A unified model of species immigration, extinction and abundance on islands. *Journal of Biogeography*, **40**, 1107-1118.

Royle, J.A. & Kéry, M. (2007) A Bayesian state-space formulation of dynamic occupancy models. *Ecology*, **88**, 1813-1823.

Rueda, M., Hawkins, B.A., Morales-Castilla, I., Vidanes, R.M., Ferrero, M. & Rodríguez, M.Á. (2013) Does fragmentation increase extinction thresholds? A European-wide test with seven forest birds. *Global Ecology and Biogeography*, **22**, 1282-1292.

Ruiz-Gutiérrez, V., Zipkin, E.F. & Dhondt, A.A. (2010) Occupancy dynamics in a tropical bird community: unexpectedly high forest use by birds classified as non-forest species. *Journal of Applied Ecology*, **47**, 621-630.

Sæther, B.-E., Engen, S. & Grøtan, V. (2013) Species diversity and community similarity in fluctuating environments: parametric approaches using species abundance distributions. *Journal of Animal Ecology*, **82**, 721-738.

Safi, K. (2011) 'Eigenvector estimation of phylogenetic and functional diversity': from patterns to processes. *Functional Ecology*, **25**, 745-746.

Sagarin, R.D. & Gaines, S.D. (2002) The 'abundant centre' distribution: to what extent is it a biogeographical rule? *Ecology Letters*, **5**, 137-147.

Sala, O.E., Chapin III , F.S., Armesto, J.J., Berlow, E., Bloomfield, J., Dirzo, R., Huber-Sanwald, E., Huenneke, L.F., Jackson, R.B., Kinzig, A., Leemans, R., Lodge, D.M., Mooney, H.A., Oesterheld, M.n., Poff, N.L., Sykes, M.T., Walker, B.H., Walker, M. & Wall, D.H. (2000) Global biodiversity scenarios for the year 2100. *Science*, **287**, 1770-1774.

Sala, O.E., Vuuren, D.v., Pereira, H.M., Lodge, D., Alder, J., Cumming, G., Dobson, A., Volters, W., Xenopoulos, M. & Zaitsev, A.S. (2005) Biodiversity across scenarios. *Ecosystem and human well-being: scenarios.* (ed. by S. Carpenter, L.P. Prabhu, E.M. Bennet and M.B. Zurek), pp. 375–408. Island Press, Washington D.C., USA.

Salisbury, C.L., Seddon, N., Cooney, C.R. & Tobias, J.A. (2012) The latitudinal gradient in dispersal constraints: ecological specialisation drives diversification in tropical birds. *Ecology Letters*, **15**, 847-855.

Sang, A., Teder, T., Helm, A. & Pärtel, M. (2010) Indirect evidence for an extinction debt of grassland butterflies half century after habitat loss. *Biological Conservation*, **143**, 1405-1413.

Santos, A.M.C., Whittaker, R.J., Triantis, K.A., Borges, P.A.V., Jones, O.R., Quicke, D.L.J. & Hortal, J. (2010) Are species–area relationships from entire archipelagos congruent with those of their constituent islands? *Global Ecology and Biogeography*, **19**, 527-540.

Sax, D.F. & Gaines, S.D. (2003) Species diversity: from global decreases to local increases. *Trends in Ecology & Evolution*, **18**, 561-566.

Scheffer, M. & van Nes, E.H. (2006) Self-organized similarity, the evolutionary emergence of groups of similar species. *Proceedings of the National Academy of Sciences USA*, **103**, 6230-6235.

Scheiner, S.M. (2003) Six types of species-area curves. *Global Ecology and Biogeography*, **12**, 441-447.

Scheiner, S.M. (2004) A mélange of curves – further dialogue about species–area relationships. *Global Ecology and Biogeography*, **13**, 479-484.

Scheiner, S.M., Chiarucci, A., Fox, G.A., Helmus, M.R., McGlinn, D.J. & Willig, M.R. (2011) The underpinnings of the relationship of species richness with space and time. *Ecological Monographs*, **81**, 195-213.
Schipper, J., Chanson, J.S., Chiozza, F. *et al.* (2008) The status of the world's land and marine mammals: diversity, threat, and knowledge. *Science*, **322**, 225-230.
Schmeller, D.S., Dolek, M., Geyer, A., Settele, J. & Brandl, R. (2011) The effect of conservation efforts on morphological asymmetry in a butterfly population. *Journal for Nature Conservation*, **19**, 161-165.
Schmidt, N.M. & Jensen, P.M. (2003) Changes in mammalian body length over 175 years—adaptations to a fragmented landscape? . *Conservation Ecology*, **7**, 6.
Schoereder, J.H., Galbiati, C., Ribas, C.R., Sobrinho, T.G., Sperber, C.F., DeSouza, O. & Lopes-Andrade, C. (2004) Should we use proportional sampling for species–area studies? *Journal of Biogeography*, **31**, 1219-1226.
Schouten, M.A., Verweij, P.A., Barendregt, A., Kleukers, R.J.M. & De Ruiter, P.C. (2007) Nested assemblages of Orthoptera species in the Netherlands: the importance of habitat features and life-history traits. *Journal of Biogeography*, **34**, 1938-1946.
Segura, A.M., Calliari, D., Kruk, C., Conde, D., Bonilla, S. & Fort, H. (2011) Emergent neutrality drives phytoplankton species coexistence. *Proceedings of the Royal Society B: Biological Sciences*, **278**, 2355-2361.
Şekercioğlu, C.H. & Sodhi, N.S. (2007) Conservation biology: predicting birds' responses to forest fragmentation. *Current Biology*, **17**, R838-R840.
Şekercioğlu, Ç.H., Ehrlich, P.R., Daily, G.C., Aygen, D., Goehring, D. & Sandí, R.F. (2002) Disappearance of insectivorous birds from tropical forest fragments. *Proceedings of the National Academy of Sciences USA*, **99**, 263-267.
Sfenthourakis, S. & Triantis, K.A. (2009) Habitat diversity, ecological requirements of species and the Small Island Effect. *Diversity and Distributions*, **15**, 131-140.
Shaffer, M.L. (1981) Minimum population sizes for species ocnservation. *BioScience*, **31**, 131-134.
Shmida, A. & Wilson, M.V. (1985) Biological determinants of species diversity. *Journal of Biogeography*, **12**, 1-20.
Simberloff, D.S. & Abele, L.G. (1976) Island biogeography theory and conservation practice. *Science*, **191**, 285-286.
Simberloff, D. & Abele, L.G. (1982) Refuge design and island biogeographic theory: effects of fragmentation. *The American Naturalist*, **120**, 41-50.
Simberloff, D. & Martin, J.L. (1991) Nestedness of insular avifaunas: simple summary statistics masking complex species patterns. *Ornis Fennica*, **68**, 178-192.
Šizling, A., Storch, D., Reif, J. & Gaston, K. (2009) Invariance in species-abundance distributions. *Theoretical Ecology*, **2**, 89-103.
Smith, A.B. (2010) Caution with curves: caveats for using the species-area relationship in conservation. *Biological Conservation*, **143**, 555-564.
Smith, A., Koper, N., Francis, C. & Fahrig, L. (2009) Confronting collinearity: comparing methods for disentangling the effects of habitat loss and fragmentation. *Landscape Ecology*, **24**, 1271-1285.
Smith, R.L. (1997) *Ecology and field biology* Fifth edn. Addison Wesley, Boston, USA.
Sólymos, P. & Lele, S.R. (2012) Global pattern and local variation in species–area relationships. *Global Ecology and Biogeography*, **21**, 109-120.
Sonnier, G., Johnson, S.E., Amatangelo, K.L., Rogers, D.A. & Waller, D.M. (2014) Is taxonomic homogenization linked to functional homogenization in temperate forests? *Global Ecology and Biogeography*, **23**, 894-902.

Soulé, M.E., Bolger, D.T., Allison, C.A., Wright, J., Sorice, M. & Hill, S. (1988) Reconstructed dynamics of rapid extinctions of chaparral-requiring birds in urban habitat islands. *Conservation Biology*, **2**, 75-92.

Southwood, T.R.E. & Kennedy, C.E.J. (1983) Trees as Islands. *Oikos*, **41**, 359-371.

Spiegelhalter, D., Thomas, A., Best, N. & Lunn, D. (2003) *WinBUGS Version 1.4 user manual*. MRC Biostatistics Unit.

Spiesman, B.J. & Inouye, B.D. (2013) Habitat loss alters the architecture of plant–pollinator interaction networks. *Ecology*, **94**, 2688-2696.

Steffan-Dewenter, I. & Tscharntke, T. (2000) Butterfly community structure in fragmented habitats. *Ecology Letters*, **3**, 449-456.

Steinbauer, M.J., Dolos, K., Reineking, B. & Beierkuhnlein, C. (2012) Current measures for distance decay in similarity of species composition are influenced by study extent and grain size. *Global Ecology and Biogeography*, **21**, 1203-1212.

Stiles, A. & Scheiner, S.M. (2010) A multi-scale analysis of fragmentation effects on remnant plant species richness in Phoenix, Arizona. *Journal of Biogeography*, **37**, 1721-1729.

Stoate, C., Boatman, N.D., Borralho, R.J., Carvalho, C.R., Snoo, G.R.d. & Eden, P. (2001) Ecological impacts of arable intensification in Europe. *Journal of Environmental Management*, **63**, 337-365.

Stohlgren, T., Chong, G., Kalkhan, M. & Schell, L. (1997) Rapid assessment of plant diversity patterns: a methodology for landscapes. *Environmental Monitoring and Assessment*, **48**, 25-43.

Stork, N. (2010) Re-assessing current extinction rates. *Biodiversity and Conservation*, **19**, 357-371.

Strona, G. & Fattorini, S. (2014) On the methods to assess significance in nestedness analyses. *Theory in Biosciences*, **133**, 1-8.

Sturtz, S., Ligges, U. & Gelman, A.E. (2005) R2WinBUGS: a package for running WinBUGS from R. *Journal of Statistical Software* **12** 1-16.

Sugihara, G. (1980) Minimal community structure: an explanation of species abundance patterns. *The American Naturalist*, **116**, 770-787.

Summerville, K.S. & Crist, T.O. (2001) Effects of experimental habitat fragmentation on patch use by butterflies and skippers (Lepidoptera). *Ecology*, **82**, 1360-1370.

Sutherland, W.J. (ed.) (2006) *Ecological census techniques: a handbook*, Second edn. Cambridge University Press, Cambridge.

Sutherland, W.J., Newton, I. & Green, R. (2004) *Bird ecology and conservation: a handbook of techniques*. Oxford University Press, Oxford.

Swift, T.L. & Hannon, S.J. (2010) Critical thresholds associated with habitat loss: a review of the concepts, evidence, and applications. *Biological Reviews*, **85**, 35-53.

Syrek, D., Weiner, W.M., Wojtylak, M., Olszowska, G. & Kwapis, Z. (2006) Species abundance distribution of collembolan communities in forest soils polluted with heavy metals. *Applied Soil Ecology*, **31**, 239-250.

Tabanez, A.A.J. & Viana, V.M. (2000) Patch structure within Brazilian Atlantic forest fragments and implications for conservation. *Biotropica*, **32**, 925-933.

Tanentzap, A.J., Walker, S., Theo Stephens, R.T. & Lee, W.G. (2012) A framework for predicting species extinction by linking population dynamics with habitat loss. *Conservation Letters*, **5**, 149-156.

Tang, H., Song, M.-Y., Cho, W.-S., Park, Y.-S. & Chon, T.-S. (2010) Species abundance distribution of benthic chironomids and other macroinvertebrates across different levels of pollution in streams. *International Journal of Limnology*, **46**, 53-66.

Taylor, L.R. (1961) Aggregation, variance and the mean. *Nature*, **189**, 732-735.

Taylor, L.R. (1986) Synoptic dynamics, migration and the Rothamsted insect survey. *Journal of Animal Ecology*, **55**, 1-38.
Tellería, J.L. & Santos, T. (1995) Effects of forest fragmentation on a guild of wintering passerines: The role of habitat selection. *Biological Conservation*, **71**, 61-67.
ter Braak, C.J.F. & Etienne, R.S. (2003) Improved Bayesian analysis of metapopulation data with an application to a tree frog metapopulation. *Ecology*, **84**, 231-241.
ter Braak, C.J.F., Hanski, I. & Verboom, J. (1997) The incidence function approach to modelling of metapopulation dynamics. *Modelling spatio-temporal dynamics in ecology* (ed. by J. Bascompte and R.V. Solé), pp. 167-188. Landes Biosciences, Austin.
Terborgh, J. (1974) Faunal equilibria and the design of wildlife preserves. Tropical ecological systems: trends in terrestrial and aquatic research (ed. by F. Golley and E. Medina), pp. 369-380. Springer Verlag, New York.
Terborgh, J., Lopez, L., Nuñez, P., Rao, M., Shahabuddin, G., Orihuela, G., Riveros, M., Ascanio, R., Adler, G.H., Lambert, T.D. & Balbas, L. (2001) Ecological meltdown in predator-free forest fragments. *Science*, **294**, 1923-1926.
Thomas, C.D. (2000) Dispersal and extinction in fragmented landscapes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **267**, 139-145.
Tilman, D. & Kareiva, P. (eds.) (1997) *Spatial ecology: the role of space in population dynamics and interspecific interactions*. Princeton University Press, Princeton.
Tilman, D., May, R.M., Lehman, C.L. & Nowak, M.A. (1994) Habitat destruction and the extinction debt. *Nature*, **371**, 65-66.
Tjørve, E. (2003) Shapes and functions of species–area curves: a review of possible models. *Journal of Biogeography*, **30**, 827-835.
Tjørve, E. (2009) Shapes and functions of species–area curves (II): a review of new models and parameterizations. *Journal of Biogeography*, **36**, 1435-1445.
Tjørve, E. & Tjørve, K.M.C. (2008) The species–area relationship, self-similarity, and the true meaning of the *z*-value. *Ecology*, **89**, 3528-3533.
Tjørve, E. & Tjørve, K.M.C. (2011) Subjecting the theory of the small-island effect to Ockham's razor. *Journal of Biogeography*, **38**, 1836-1839.
Tobias, J.A., Cornwallis, C.K., Derryberry, E.P., Claramunt, S., Brumfield, R.T. & Seddon, N. (2014) Species coexistence and the dynamics of phenotypic evolution in adaptive radiation. *Nature*, **506**, 359-363.
Tokeshi, M. (1990) Niche apportionment or random assortment: species abundance patterns revisited. *Journal of Animal Ecology*, **59**, 1129-1146.
Tokeshi, M. (1993) Species abundance patterns and community structure. *Advances in Ecological Research*, **24**, 111-186.
Tokeshi, M. (1999) *Species coexistence: ecological and evolutionary perspectives*. Blackwell Science, Oxford.
Toms, J.D. & Lesperance, M.L. (2003) Piecewise regression: a tool for identifying ecological thresholds. *Ecology*, **84**, 2034-2041.
Traill, L.W., Bradshaw, C.J.A. & Brook, B.W. (2007) Minimum viable population size: a meta-analysis of 30 years of published estimates. *Biological Conservation*, **139**, 159-166.
Triantis, K.A. & Bhagwat, S.A. (2011) Applied island biogeography. *Conservation biogeography* (ed. by R. Ladle and R.J. Whittaker). Wiley-Blackwell, Oxford.
Triantis, K.A. & Sfenthourakis, S. (2012) Island biogeography is not a single-variable discipline: the small island effect debate. *Diversity and Distributions*, **18**, 92-96.
Triantis, K.A., Mylonas, M., Lika, K. & Vardinoyannis, K. (2003) A model for the species-area-habitat relationship. *Journal of Biogeography*, **30**, 19-27.

Triantis, K.A., Vardinoyannis, K., Tsolaki, E.P., Botsaris, I., Lika, K. & Mylonas, M. (2006) Re-approaching the small island effect. *Journal of Biogeography*, **33**, 914-923.
Triantis, K.A., Borges, P.A.V., Ladle, R.J., Hortal, J., Cardoso, P., Gaspar, C., Dinis, F., Mendonça, E., Silveira, L.M.A., Gabriel, R., Melo, C., Santos, A.M.C., Amorim, I.R., Ribeiro, S.P., Serrano, A.R.M., Quartau, J.A. & Whittaker, R.J. (2010) Extinction debt on oceanic islands. *Ecography*, **33**, 285-294.
Triantis, K.A., Guilhaumon, F. & Whittaker, R.J. (2012) The island species–area relationship: biology and statistics. *Journal of Biogeography*, **39**, 215-231.
Trisos, C.H., Petchey, O.L. & Tobias, J.A. (2014) Unraveling the interplay of community assembly processes acting on multiple niche axes across spatial scales. *The American Naturalist*, **184**, 593-608.
Tscharntke, T., Şekercioğlu, C.H., Dietsch, T.V., Sodhi, N.S., Hoehn, P. & Tylianakis, J.M. (2008) Landscape constraints on functional diversity of birds and insects in tropical agroecosystems. *Ecology*, **89**, 944-951.
Ugland, K.I. & Gray, J.S. (1982) Lognormal distributions and the concept of community equilibrium. *Oikos*, **39**, 171-178.
Ugland, K.I., Gray, J.S. & Ellingsen, K.E. (2003) The species–accumulation curve and estimation of species richness. *Journal of Animal Ecology*, **72**, 888-897.
Ugland, K.I., Lambshead, P.J.D., McGill, B., Gray, J.S., O'Dea, N., Ladle, R.J. & Whittaker, R.J. (2007) Modelling dimensionality in species abundance distributions: description and evaluation of the Gambin model. *Evolutionary Ecology Research*, **9**, 313-324.
Ulrich, W. & Ollik, M. (2004) Frequent and occasional species and the shape of relative-abundance distributions. *Diversity and Distributions*, **10**, 263-269.
Ulrich, W. & Zalewski, M. (2006) Abundance and co-occurrence patterns of core and satellite species of ground beetles on small lake islands. *Oikos*, **114**, 338-348.
Ulrich, W. & Gotelli, N.J. (2007) Null model analysis of species nestedness patterns. *Ecology*, **88**, 1824-1831.
Ulrich, W. & Gotelli, N.J. (2010) Null model analysis of species associations using abundance data. *Ecology*, **91**, 3384-3397.
Ulrich, W. & Gotelli, N.J. (2012) A null model algorithm for presence–absence matrices based on proportional resampling. *Ecological Modelling*, **244**, 20-27.
Ulrich, W. & Almeida-Neto, M. (2012) On the meanings of nestedness: back to the basics. *Ecography*, **35**, 865-871.
Ulrich, W. & Gotelli, N.J. (2013) Pattern detection in null model analysis. *Oikos*, **122**, 2-18.
Ulrich, W., Almeida-Neto, M. & Gotelli, N.J. (2009) A consumer's guide to nestedness analysis. *Oikos*, **118**, 3-17.
Ulrich, W., Ollik, M. & Ugland, K.I. (2010) A meta-analysis of species–abundance distributions. *Oikos*, **119**, 1149-1155.
Unterseher, M., Jumpponen, A.R.I., ÖPik, M., Tedersoo, L., Moora, M., Dormann, C.F. & Schnittler, M. (2011) Species abundance distributions and richness estimations in fungal metagenomics – lessons learned from community ecology. *Molecular Ecology*, **20**, 275-285.
Van Der Mescht, L., Le Roux, P.C. & Matthee, S. (2013) Remnant fragments within an agricultural matrix enhance conditions for a rodent host and its fleas. *Parasitology*, **140**, 368-377.
van Dorp, D. & Opdam, P.F.M. (1987) Effects of patch size, isolation and regional abundance on forest bird communities. *Landscape Ecology*, **1**, 59-73.
van Vuuren, D.P., Sala, O.E. & Pereira, H.M. (2006) The future of vascular plant diversity under four global scenarios. *Ecology and Society*, **11**, 25-42.

Veech, J.A., Summerville, K.S., Crist, T.O. & Gering, J.C. (2002) The additive partitioning of species diversity: recent revival of an old idea. *Oikos*, **99**, 3-9.
Vergnon, R., van Nes, E.H. & Scheffer, M. (2012) Emergent neutrality leads to multimodal species abundance distributions. *Nature Communications*, **3**, 663.
Villard, M.-A. & Metzger, J.P. (2014) Beyond the fragmentation debate: a conceptual model to predict when habitat configuration really matters. *Journal of Applied Ecology*, **51**, 309-318.
Villéger, S., Mason, N.W.H. & Mouillot, D. (2008) New multidimensional functional diversity indices for a multifaceted framework in functional ecology. *Ecology*, **89**, 2290-2301.
Volkov, I., Banavar, J.R., Hubbell, S.P. & Maritan, A. (2003) Neutral theory and relative species abundance in ecology. *Nature*, **424**, 1035-1037.
Wagner, C.E., Harmon, L.J. & Seehausen, O. (2014) Cichlid species-area relationships are shaped by adaptive radiations that scale with area. *Ecology Letters*, **17**, 583-592.
Wang, Y., Chen, S. & Ding, P. (2011) Testing multiple assembly rule models in avian communities on islands of an inundated lake, Zhejiang Province, China. *Journal of Biogeography*, **38**, 1330-1344.
Wang, Y., Ding, P., Chen, S. & Zheng, G. (2013) Nestedness of bird assemblages on urban woodlots: Implications for conservation. *Landscape and Urban Planning*, **111**, 59-67.
Wang, Y., Bao, Y., Yu, M., Xu, G. & Ding, P. (2010) Nestedness for different reasons: the distributions of birds, lizards and small mammals on islands of an inundated lake. *Diversity and Distributions*, **16**, 862-873.
Warwick, R.M. (1986) A new method for detecting pollution effects on marine macrobenthic communities. *Marine Biology*, **92**, 557-562.
Watling, J.I. & Donnelly, M.A. (2006) Fragments as islands: a synthesis of faunal responses to habitat patchiness. *Conservation Biology*, **20**, 1016-1025.
Watling, J.I. & Donnelly, M.A. (2008) Species richness and composition of amphibians and reptiles in a fragmented forest landscape in northeastern Bolivia. *Basic and Applied Ecology*, **9**, 523-532.
Watling, J.I., Nowakowski, A.J., Donnelly, M.A. & Orrock, J.L. (2011) Meta-analysis reveals the importance of matrix composition for animals in fragmented habitat. *Global Ecology and Biogeography*, **20**, 209-217.
Watson, D.M. (2002) A conceptual framework for studying species composition in fragments, islands and other patchy ecosystems. *Journal of Biogeography*, **29**, 823-834.
Watson, D.M. (2003) Long-term consequences of habitat fragmentation–highland birds in Oaxaca, Mexico. *Biological Conservation*, **111**, 283-303.
Watson, J., Freudenberger, D. & Paull, D. (2001) An assessment of the focal-species approach for conserving birds in variegated landscapes in southeastern Australia. *Conservation Biology*, **15**, 1364-1373.
Watson, J.E.M., Whittaker, R.J. & Freudenberger, D. (2005) Bird community responses to habitat fragmentation: how consistent are they across landscapes? *Journal of Biogeography*, **32**, 1353-1370.
Watson, J.E.M., Grantham, H.S., Wilson, K.A. & Possingham, H.P. (2011) Systematic conservation planning: past, present and future. *Conservation biogeography* (ed. by R.J. Ladle and R.J. Whittaker), pp. 136-160. Wiley-Blackwell, Chichester.
Watson, J.E.M., Watson, A.W.T., Fischer, J., Ingram, J.C. & Whittaker, R.J. (2009) Using nestedness and species-accumulation analyses to strengthen a conservation plan for littoral forest birds in south-eastern Madagascar. *International Journal of Biodiversity and Conservation*, **1**, 67-80.

Wethered, R. & Lawes, M.J. (2005) Nestedness of bird assemblages in fragmented Afromontane forest: the effect of plantation forestry in the matrix. *Biological Conservation*, **123**, 125-137.
Whitcomb, R.F., Robbins, C.S., Lynch, J.F., Whitcomb, B.L., Klimkiewicz, M.K. & Bystrak, D. (1981) Effects of forest fragmentation on avifauna of the Eastern Deciduous Forest. *Forest island dynamics in man-dominated landscapes* (ed. by R.L. Burgess and D.M. Sharpe), pp. 125-205. Springer-Verlag, New York.
White, E.P., Thibault, K.M. & Xiao, X. (2012) Characterizing species abundance distributions across taxa and ecosystems using a simple maximum entropy model. *Ecology*, **93**, 1772-1778.
Whittaker, R.H. (1965) Dominance and diversity in land plant communities. *Science*, **147**, 250–260.
Whittaker, R.H. (1967) Gradient analysis of vegetation. *Biological Reviews*, **42**, 207-264.
Whittaker, R.J. (2010) Meta-analyses and mega-mistakes: calling time on meta-analysis of the species richness–productivity relationship. *Ecology*, **91**, 2522-2533.
Whittaker, R.J. & Fernández-Palacios, J.M. (2007) *Island biogeography: ecology, evolution, and conservation*, 2nd edn. Oxford University Press, Oxford.
Whittaker, R.J. & Triantis, K.A. (2012) The species–area relationship: an exploration of that 'most general, yet protean pattern'1. *Journal of Biogeography*, **39**, 623-626.
Whittaker, R.J. & Matthews, T.J. (2014) The varied form of species–area relationships. *Journal of Biogeography*, **41**, 209-210.
Whittaker, R.J., Araújo, M.B., Jepson, P., Ladle, R.J., Watson, J.E.M. & Willis, K.J. (2005) Conservation biogeography: assessment and prospect. *Diversity and Distributions*, **11**, 3-23.
Whittaker, R.J., Rigal, F., Borges, P.A.V., Cardoso, P., Terzopoulou, S., Casanoves, F., Pla, L., Guilhaumon, F., Ladle, R.J. & Triantis, K. (2014) Functional biogeography of oceanic islands and the scaling of functional diversity in the Azores. *Proceedings of the National Academy of Sciences USA*, **111**, 13709–13714.
Wickham, H. (2009) *ggplot2: elegant graphics for data analysis*. Springer, New York.
Wiens, J.A. (1976) Population responses to patchy environments. *Annual Review of Ecology and Systematics*, **7**, 81-120.
Wiens, J.A. (1995) Habitat fragmentation: island v landscape perspectives on bird conservation. *Ibis*, **137**, S97-S104.
Wiens, J.A. & Milne, B.T. (1989) Scaling of 'landscapes' in landscape ecology, or, landscape ecology from a beetle's perspective. *Landscape Ecology*, **3**, 87-96.
Wilcove, D.S. (1987) From fragmentation to extinction. *Natural Areas Journal*, **7**, 23-29.
Wilcove, D.S., C.H, M. & A.P, D. (1986) Habitat fragmentation in the temperate zone. *Conservation biology: the science of scarcity and diversity* (ed. by M.E. Soulé), pp. 237–256. Sinauer Associates, Sunderland.
Wilcox, B.A. (1978) Supersaturated island faunas: a species-age relationship for lizards on post-Pleistocene land-bridge islands. *Science*, **199**, 996-998.
Wilcox, B.A. & Murphy, D.D. (1985) Conservation strategy: the effects of fragmentation on extinction. *The American Naturalist*, **125**, 879-887.
Williams, C.B. (1964) *Patterns in the balance of nature*. Academic Press, London.
Williams, M.R., Lamont, B.B. & Henstridge, J.D. (2009) Species–area functions revisited. *Journal of Biogeography*, **36**, 1994-2004.
Williamson, M. (1981) *Island populations*. Oxford University Press, Oxford.
Williamson, M. & J. Gaston, K. (2005) The lognormal distribution is not an appropriate null hypothesis for the species–abundance distribution. *Journal of Animal Ecology*, **74**, 409-422.

Williamson, M., Gaston, K.J. & Lonsdale, W.M. (2001) The species–area relationship does not have an asymptote! *Journal of Biogeography*, **28**, 827-830.
Willis, K.J. & Whittaker, R.J. (2002) Species diversity: scale matters. *Science*, **295**, 1245-1248.
Wilson, E.O. & Willis, E.O. (1975) Applied biogeography. *Ecology and evolution of communities* (ed. by M.L. Cody and J.M. Diamond), pp. 522-534. Harvard University Press, Cambridge, MA.
Wilson, J.B. (2007) Priorities in statistics, the sensitive feet of elephants, and don't transform data. *Folia Geobotanica*, **42**, 161-167.
Wilson, M.V. & Shmida, A. (1984) Measuring beta diversity with presence–absence data. *Journal of Ecology*, **72**, 1055-1064.
With, K.A. & Crist, T.O. (1995) Critical thresholds in species responses to landscape structure. *Ecology*, **76**, 2446-2459.
With, K.A. & King, A.W. (1999) Dispersal success on fractal landscapes: a consequence of lacunarity thresholds. *Landscape Ecology*, **14**, 73-82.
With, K.A., Pavuk, D.M., Worchuck, J.L., Oates, R.K. & Fisher, J.L. (2002) Threshold effects of landscape structure on biological control in agroecosystems. *Ecological Applications*, **12**, 52-65.
World Weather Online (2014) *Rakkestad weather, Norway*. Available at: http://www.worldweatheronline.com/ (accessed August 2014).
Wright, D.H. (1983) Species-energy theory: an extension of species-area theory. *Oikos*, **41**, 496-506.
Wright, D.H. (1991) Correlations between incidence and abundance are expected by chance. *Journal of Biogeography*, **18**, 463-466.
Wright, D.H., Patterson, B.D., Mikkelson, G.M., Cutler, A. & Atmar, W. (1998) A comparative analysis of nested subset patterns of species composition. *Oecologia*, **113**, 1-20.
Yaacobi, G., Ziv, Y. & Rosenzweig, M.L. (2007) Habitat fragmentation may not matter to species diversity. *Proceedings of the Royal Society B: Biological Sciences*, **274**, 2409-2412.
Yong, D.L., Qie, L., Sodhi, N.S., Koh, L.P., Peh, K.S.-H., Lee, T.M., Lim, H.C. & Lim, S.L.-H. (2011) Do insectivorous bird communities decline on land-bridge forest islands in peninsular Malaysia? *Journal of Tropical Ecology*, **27**, 1-14.
Zimmerman, B.L. & Bierregaard, R.O. (1986) Relevance of the equilibrium-theory of island biogeography and species area relations to conservation with a case from amazonia. *Journal of Biogeography*, **13**, 133-143.
Zuckerberg, B. & Porter, W.F. (2010) Thresholds in the long-term responses of breeding birds to forest cover and fragmentation. *Biological Conservation*, **143**, 952-962.
Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A. & Smith, G.M. (2009) *Mixed effects models and extensions in ecology with R*. Springer, New York.

Appendix 1

Neutral theory and the species abundance distribution: recent developments and prospects for unifying niche and neutral perspectives

Paper published in Ecology & Evolution: Matthews, T.J. & Whittaker, R.J. (2014) Neutral theory and the species abundance distribution: recent developments and prospects for unifying niche and neutral perspectives. *Ecology & Evolution*, **4,** 2263–2277.

A1.1 Preface

This appendix reviews the development of ecological neutral theory, with a particular focus on the species abundance distribution. It also reviews recent developments aimed at unifying neutral and niche based approaches to ecology.

Authors: Thomas J. Matthews and Robert J. Whittaker

The co-author contributions were as follows. T.J.M. conceived the ideas; and T.J.M. and R.J.W. wrote the paper.

A1.2 Abstract

Published in 2001, The Unified Neutral Theory of Biodiversity and Biogeography (UNTB) emphasises the importance of stochastic processes in ecological community structure, and has challenged the traditional niche-based view of ecology. While neutral models have since been applied to a broad range of ecological and macroecological phenomena, the majority of research relating to neutral theory has focused exclusively on the species abundance distribution (SAD). Here, we synthesise the large body of work on neutral theory in the context of the species abundance distribution, with a particular focus on integrating ideas from neutral theory with traditional niche theory. First, we summarise the basic tenets of neutral theory; both in general and in the context of SADs. Second, we explore the issues associated with neutral theory and the SAD, such as complications with fitting and model comparison, the underlying assumptions of neutral models, and the difficultly of linking pattern to process. Third, we highlight the advances in understanding of SADs that have resulted from neutral theory and models. Finally, we focus consideration on recent developments aimed at unifying neutral and niche based approaches to ecology, with a particular emphasis on what this means for SAD theory, embracing, for instance, ideas of emergent neutrality and stochastic niche theory. We put forward the argument that the prospect of the unification of niche and neutral perspectives represents one of the most promising future avenues of neutral theory research.

A1.3 A brief history of neutral ecology

The dominant view in ecology in the first half of the twentieth century was that of stability and of ecological communities structured through mechanisms such as competitive interactions, density-dependence, local adaptations and niche differentiation (Begon *et al.,* 1990). This paradigm was challenged through various theories published in the mid-twentieth century, such as MacArthur & Wilson's (1967) equilibrium theory of island biogeography

(ETIB), with researchers looking at broader-scale structuring mechanisms. More recently Hubbell (2001) generated substantial debate when he challenged the traditional niche-based view of community structure by claiming that numerous macroecological patterns, including species abundance distributions (SADs; see Table A1.1 for definitions of key terms and acronyms used in this paper) and species–area relationships (SARs), could be explained through his Unified Neutral Theory of Biodiversity and Biogeography (UNTB). Following Caswell (1976), Hubbell (2001, 2006) put forward the argument that species differences are not necessarily evidence of important roles for niche differentiation, criticising the 'uncritical acceptance' of such evidence in the ecological literature. Thus, neutral theory rejects two concepts that have formed the basis of traditional ecological research and niche theory (Chesson, 1991), namely that species are ecologically and functionally different, and that environmental context is important (McGill, 2006; Rosindell & Cornell, 2007). In this regard neutral theory is frequently misinterpreted. It does not imply that species are undifferentiated in their traits. Rather it asks the questions: do these differences matter in terms of ecological community structure, and are they functionally significant?

Hubbell (2001) evaluated the UNTB by means of an extensive analysis of a dataset from Barro Colorado Island (herein 'BCI'). This dataset is a count of all tree and shrub species greater than one cm in stem diameter from a 50ha plot. He found that by focusing on all species it was possible to observe significant differences between species, implying niche differentiation, but that 75% of the species occupied a similar shade tolerant niche. Hubbell (2001; 2006) argued that niche-based theories did not provide explanations for the multitude of such species; it being more likely that such a pattern arose through the abundance of shady habitat compared to open habitats over the evolutionary history of the species. Thus, these species have simply evolved life histories adapted to shady habitats regardless of other species life histories. This has led to the coexistence of ecologically equivalent species

(Hubbell, 2006), a pattern seemingly explained through the UNTB. Furthermore, as individuals in species rich communities frequently have numerous different nearest neighbours relative to conspecific individuals (an argument constructed by reference to plants) any local adaptation is unlikely to be along the same direction between individuals. This, it was argued, led not to niche divergence and specialist life histories but to niche convergence towards generalist strategies and, therefore, ecological equivalence (Hubbell & Foster, 1986; Hubbell ,2006).

Neutrality was not a new concept, having been applied in the field of population genetics during the 1960s (e.g. Kimura & Crowm, 1964; Kimura, 1968; King & Jukes, 1969; Watterson 1974a; for a review see Leigh, 2007). In this context neutrality refers to a situation where there is equal likelihood that a gene enters the next generation irrespective of its allelic type. The theory was subsequently applied to ecology (e.g. Watterson, 1974b; Caswell, 1976). However, whereas these previous applications of neutral theory had considered the importance of random genetic drift, Hubbell also postulated that random dispersal was a primary controlling factor of community structure, i.e. the UNTB is a dispersal-assembled theory (Hubbell, 2001). The UNTB asserts that species abundances result from a combination of dispersal, speciation, and stochastic variation in birth and death rates (ecological drift). Thus, the abundance of species changes by chance and not because of differences in competitive ability (Hubbell, 2001; Etienne & Olff, 2004). Extinction is also an important process but neutral models assume that extinction rate can be predicted if values of the other processes are known (Borda-de-Água *et al.,* 2007). In Hubbell's (2001) original work, dispersal was modelled in various ways. For instance, in chapter 5 dispersal was based on the 'voter model', in which species can only disperse to sites immediately adjacent to those they occupy at each given time step. Such a model is unrealistic and subsequent work (Borda-de-Água *et al.,* 2007; see also Rosindell & Cornell, 2013) has advanced this approach to

incorporate dispersal kernels ('Lévy flights') that allow for long-distance dispersal. Neutral theory also assumes that as species are functionally equivalent, the rates of these processes are the same for each species on a per capita basis and that variation in abundance between species is a result of 'accidental' dispersal and ecological drift (Hubbell, 2001; Holt, 2006; Borda-de-Água *et al.*, 2007). Hubbell (2001) originally contended that the theory should be applicable to multiple taxa; albeit only when considering trophically similar species.

Neutrality can be condensed into two forms: hard and weak. Weak neutrality describes the situation in which null models moderately explain community structure, even if fine scale examination reveals that species do differ in their niche characteristics. Hard neutrality relates to a situation whereby species are completely functionally equivalent and species identity does not matter in explaining macroecological patterns, such as SADs (Bell, 2001; Holt, 2006). Furthermore, around ten different variants of neutral models have been discussed in the ecological literature, each with slightly different predictions for different factors. Hubbell's (2001) neutral model is a spatially implicit neutral model (SINM) that focuses on two scales of community organisation: the metacommunity (source pool of individuals) and the local community. This model is also referred to as the mainland–island model, where the metacommunity represents the mainland source pool of individuals with the potential to disperse to the island community. Recent work by Rosindell and colleagues (e.g. Rosindell *et al.*, 2008; Rosindell & Cornell, 2013) has developed spatially explicit neutral models (SENM) in which the two-tier community hierarchy (meta- and local community) is replaced with a multi-scale structure. While authors often use the term 'neutral theory' to describe particular models (e.g. Hubbell's SINM), we agree with Rosindell *et al.* (2012) that the term should refer to the larger collection of neutral models.

Although the present review is focused largely on SADs, the application of neutral theory is not restricted to SADs, and numerous macroecological patterns have been explored

using neutral models, such as species–area relationships (SARs), abundance–occupancy relationships, species turnover and the distance-decay relationship (Bell, 2001; Chave & Leigh, 2002). This conceptual breadth is one of the advantages of the theory.

A number of other reviews of neutral theory exist (e.g. Alonso *et al.*, 2006; Leigh, 2007; Rosindell *et al.*, 2011). However, these approach neutral theory in a more general way than our review. Here our purpose is to provide an in-depth analysis of the species abundance distribution in the context of neutral theory. Furthermore, past reviews have focused solely on neutral theory proper (e.g. Rosindell *et al.*, 2011), rather than on the idea of unifying neutral and niche perspectives, and as such do not cover recent advancements such as stochastic niche theory and emergent neutrality. A considerable amount of research has been undertaken on integrating aspects of niche and neutral theories in the last ten years and we feel a review of this material may provide a useful resource which a) condenses the considerable amount work that has already been published, and b) provides a catalyst and guide for future work in this direction.

A1.4 Neutral theory and the species abundance distribution

A multivariate species abundance distribution (herein 'SAD') describes the abundances of all species sampled within a given community (Table A1.1; Alonso *et al.*, 2008; Ulrich *et al.*, 2010). To date around 30 different SAD models have been published (McGill *et al.*, 2007). The two most commonly used are the logseries (Fisher *et al.*, 1943) and the lognormal (Preston 1948). The logseries distribution results from the Poisson sampling of a gamma distribution: the modal abundance value in a logseries distribution is one (Fisher *et al.*, 1943). The lognormal distribution represents a situation in which the logarithms of the different species' abundances follow a Gaussian distribution, and as such it characterises a community with relatively few very abundant or very rare species.

Table A1.1 Neutral theory and species-abundance distribution terminology.

Term	Definition
Emergent neutrality (EN)	A model in which neutrality is the outcome of community evolution. According to the model competing species self-organise into groups of species with similar traits.
Fundamental biodiversity number	A dimensionless parameter in neutral models which describes various characteristics of the metacommunity. The parameter (θ) is given by $\mathrm{Jm}\frac{v}{1-v}$, where Jm is the number of individuals in the metacommunity, and v is the probability that a deceased individual is replaced by a new species through speciation (per capita speciation rate).
Fundamental immigration number	A parameter (I) given by $\frac{m(\mathrm{J}-1)}{1-m}$, where J is the number of individuals in the local community and m is the probability that a deceased individual in the local community is replaced by an immigrant from the metacommunity. It is often used in neutral models as a measure of dispersal limitation as the parameter m can be difficult to interpret. 'I' is also independent of sample size and can be seen as measure of community isolation.
Likelihood surface	The value of the likelihood, usually displayed in graphical form, as a function of a number of parameters (generally two or three). Multiple local maxima refer to the situation in which the likelihood surface has more than one peak.
Local community	Generally relates to the local community in Hubbell's spatially implicit neutral model, in which it is a set of individuals that live in the same smaller sample/island distinct from the larger metacommunity/mainland. A dead individual is immediately replaced either by an offspring of another individual (of any species) in the local community, or by an immigrant from the metacommunity (determined by m; the probability that a deceased individual in the local community is replaced by an immigrant from the metacommunity). As long as m is greater than 0 the local community receives a certain amount of immigrants from the metacommunity. The number of individuals in the local community is predicted to be too small for speciation to occur.
Lognormal distribution	A probability distribution of a random variable whose logarithm follows a Gaussian distribution. In relation to species abundance distributions, the lognormal distribution characterises a sample with relatively few very abundant or very rare species.
Logseries distribution	A probability distribution which results from the Poisson sampling of a gamma distribution after a certain relevant limit is taken, and conditional presence is considered, i.e., it gives the conditional probability of attaining a certain abundance level given that the species is present . In relation to species abundance distributions, the logseries distribution characterises a sample in which the most common abundance category is a single individual.
Metacommunity	In the context of neutral theory the metacommunity generally relates to Hubbell's spatially implicit neutral model, in which it is the source pool of

	individuals. A dead individual is immediately replaced either by an offspring of another organism in the metacommunity, or by an individual from a new species (speciation). Offspring of individuals may disperse to the local community (above).
Multimodal species abundance distribution	A species abundance distribution with multiple modal abundance values or octaves. The majority of published species abundance distribution models are unimodal, but it has become increasingly apparent that many empirical abundance distributions exhibit multiple modes
Spatially explicit neutral model (SENM)	A neutral model that incorporates an explicit spatial structure, which enables the model to predict the exact location of each individual in space.
Spatially implicit neutral model (SINM)	A neutral model that incorporates a restricted consideration of spatial structure. Hubbell's (2001) classic neutral model is spatially implicit as it only focuses on two scales of community organisation, i.e. the metacommunity and the local community.
Speciation mode (within neutral models)	The manner in which speciation is modelled in neutral models. In Hubbell's classic SINM, speciation occurs via the point mutation mode whereby speciation is an instantaneous process. Neutral models incorporating alternative speciation modes have since been developed; for instance, whereby speciation is a gradual, drawn out process (protracted speciation).
Species abundance distribution (SAD)	The typical univariate SAD gives the expected frequency of species at each abundance level, either in terms of relative frequencies or simply by the average number of species at each abundance level. The multivariate SAD gives the whole multidimensional distribution: the abundance of all species observed within a sample of an ecological community.
Species–area relationship (SAR)	The relationship between the area of a sample or island and of the number of species in that area.
Stochastic niche theory	A theory of community structure which combines niche apportionment with stochastic processes
Zero-sum assumption	An assumption of many neutral models, including Hubbell's (2001) SINM, stipulating that when an individual dies it is immediately replaced by another individual, i.e. resources are fully saturated at all times.
Zero-sum multinomial distribution (ZSM)	The species abundance distribution predicted for the local community in Hubbell's (2001) SINM (above).

Hubbell's (2001) general argument in relation to SADs was that the lognormal gives an underestimate of the number of rare species in species rich communities; that is, it did not deal with the log-left skew observed in many natural systems (Gray *et al.*, 2006). In Hubbell's SINM the relative abundance of species within – and the species diversity of – a

community can be explained through neutral drift of individual species' abundances. The model contends that the number of individuals in a metacommunity (J_m) is constant, i.e. all available resources in the community are saturated. This is the zero-sum assumption: if an individual dies and a portion of the resource becomes available, it will be immediately taken up by a new individual, and the community size remains constant (Hubbell, 2001). Certain studies have relaxed the zero-sum assumption (e.g. Volkov *et al.,* 2003; Etienne *et al.,* 2007a; see discussion in Alonso *et al.,* 2006) but it has been found that this does not change the form of the SAD (Etienne *et al.,* 2007a). The deceased individual is replaced either via a new species through speciation (with probability *v*) or via the offspring of a randomly selected individual from a species already present in the community (probability 1-*v*; see Fig. A1.1 for an illustration of this process). The assumption of neutrality implies that these probabilities (i.e. that an individual undergoes speciation, or reproduces) are identical for all species in the community on a per capita basis. The SAD can be determined using what Hubbell termed the 'fundamental biodiversity number' (θ), whereby $\theta = J_m \frac{v}{1-v}$, or in the case of non-overlapping generations $\theta = 2J_m v$. This dimensionless biodiversity parameter is independent of sample size, specifying the number of new species that appear in the community on average per generation, the number of predicted species present in the metacommunity at a steady-state between extinction and speciation, and the predicted abundance of each species at metacommunity scale. The parameter is also proportional to both the size of the metacommunity and the average per capita metacommunity speciation rate (Hubbell, 2001; He & Hu, 2005; Hubbell *et al.,* 2008). It has also transpired that θ is equivalent to Fisher's α and the expected SAD for the metacommunity is Fisher *et al.*'s (1943) logseries (described above; see also Watterson, 1974b; Hubbell *et al.,* 2008). The parameter θ has also been found to have an analytical relationship with Simpson's diversity index (He & Hu, 2005).

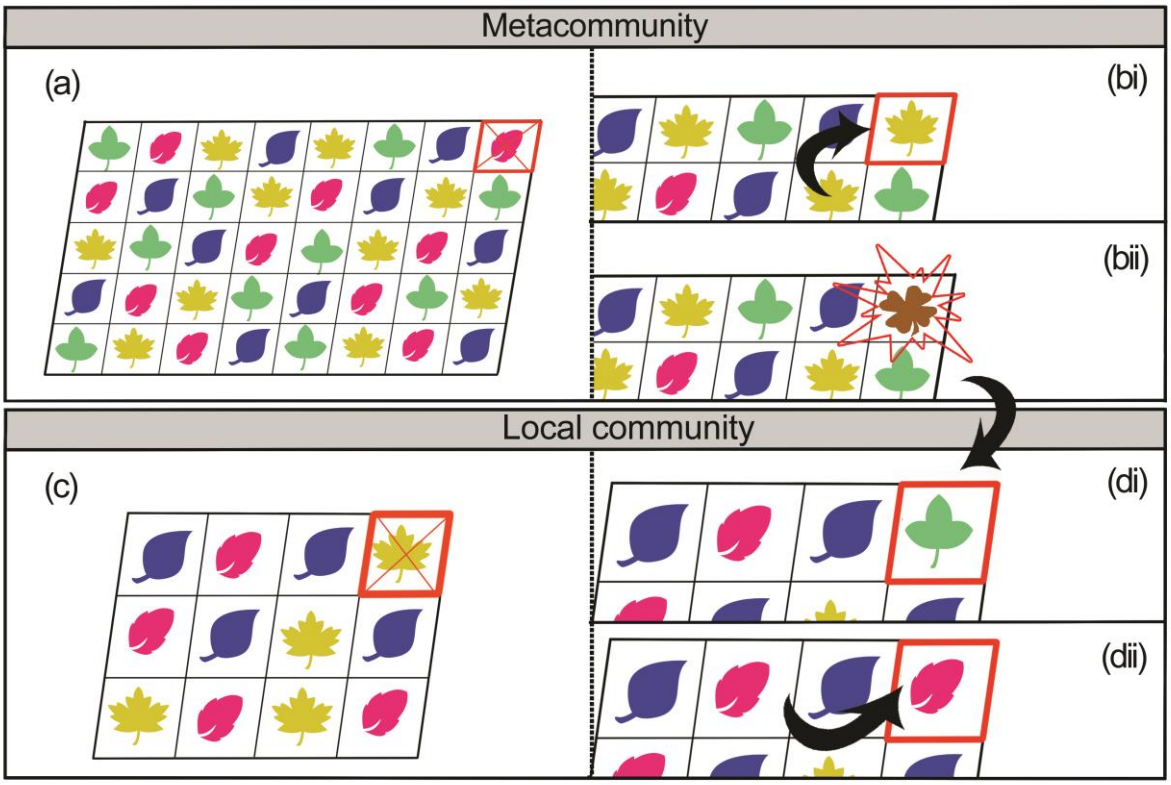


Figure A1.1 An illustration of Hubbell's (2001) classic two-tier spatially implicit neutral model. The different coloured leaves represent different species of trees. The top row (a & b) represents the metacommunity (a large continuous forest which is the source pool of individuals) and the bottom row (c & d) represents the local community (a smaller distinct patch of forest). In the metacommunity (a), at time step t a random individual (highlighted by a red square and cross) dies and is instantly replaced at time step t + 1 (the zero-sum assumption) by either the offspring of another individual in the community (bi; with probability 1-*v*) or through the instantaneous emergence (point mutation) of a new species, i.e. speciation occurs (bii; with probability *v*). A similar process occurs in the local community, but here immigration replaces speciation. At time step t a randomly chosen individual dies (c; highlighted by a red square and cross) and is instantly replaced by either an immigrant from the metacommunity (di; with probability *m*), or via the offspring of any species in the local community (dii; with probability 1-*m*), at t + 1. The local community is generally assumed to be panmictic, i.e. dispersal limitation is ignored and any individual has the same probability of producing the offspring that replaces the deceased individual.

In the SINM the local community is embedded within the wider metacommunity, with the former often provided with immigrants from the latter. At the local community scale (population size J) there is no speciation and, at the death of an individual, it is replaced either by a randomly selected immigrant from the metacommunity (probability of *m*) or via the offspring of a random individual of a species already present within the local community

(probability of 1-*m*; see Fig. A1.1). If there is no immigration into the local community from the metacommunity (i.e. m=0), then the SINM predicts that eventually only one species will remain (termed 'monodominance'), as all others are lost through a random walk to extinction. The 'fundamental dispersal number (I),' calculated as $I = \frac{m(J-1)}{1-m}$ (for non-overlapping generations $I=2Jm$), is often used because interpreting m as a measure of dispersal limitation is problematic. Hubbell's (2001) SINM predicts that the SAD at the local community level will follow a zero-sum multinomial distribution (ZSM; also termed the dispersal limited multinomial by later studies; e.g. Etienne & Alonso 2005). The ZSM is parameterised by θ, m and J (Hubbell, 2001). The parameter m is central in the shaping of the SAD of the local community and as m approaches zero (low migration communities) the local community supports fewer rarer species as it progresses towards monodominance. For a given value of θ a high migration system (m>0.1) will possess higher local diversity than a low migration system (m<0.1); the latter always having higher turnover (Latimer *et al.*, 2005). At m=0 evidently only one species will be present, but at intermediate values of m the local community SAD will have a lognormal type shape. Thus, shifts in the value of m offer an explanation for the log-left skew of the standard lognormal form that has been observed in many ecological studies (i.e. there are often more rare species than predicted by the lognormal; see Fig. A1.2 for an example of the fitted ZSM), and indeed the prevalence of logseries distributions (Bell, 2001; Magurran, 2005). The ability of the ZSM to unite the shapes of the lognormal and logseries distributions results in the good fit to empirical data. Nonetheless, it has been noted that m and θ can also be seen as purely 'geometric descriptors' of the shape of the distribution without any assumption of the validity of the neutral model that underpins them (McGill *et al.*, 2006; McGill, 2011). There is also an issue in using maximum likelihood methods to estimate the parameter values, as while this approach may

produce parameter values that result in a good statistical fit, the estimated values may be ecologically unrealistic (Gotelli & McGill, 2006; McGill *et al.*, 2006).

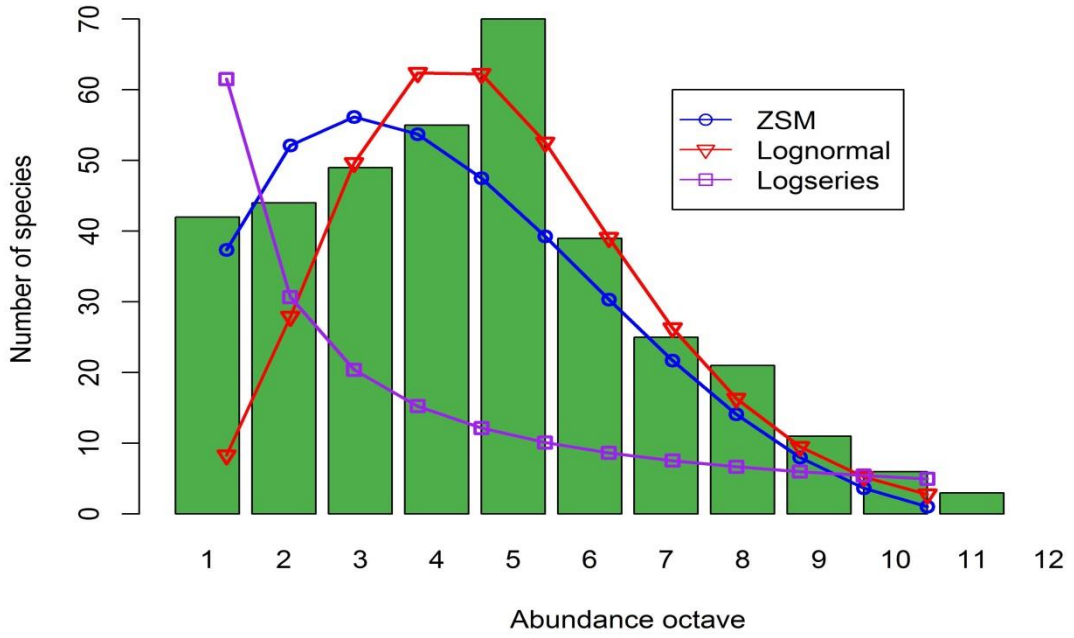


Figure A1.2 Exemplar fits of three species abundance distribution models: the zero-sum multinomial distribution of Hubbell's (2001) spatially implicit neutral model, the Poisson lognormal distribution (Preston, 1948), and the logseries distribution (Fisher *et al.*, 1943). The models are fitted to simulated data (green bars; 365 species and 22945 individuals). The three models are fitted using maximum likelihood methods. The simulated data are binned into octaves following method 3 in Gray *et al.* (2006): the first octave contains the number of species represented one individual, the second octave contains the number of species with 2-3 individuals, the third octave represents 4-7 individuals, and so on. The asymmetry of the ZSM enables it to provide a better fit than the other distributions to the left hand tail of the empirical distribution.

The shape of the SAD has also been explored using SENMs. For instance, Rosindell & Cornell (2013) investigated the SADs predicted by a SENM at multiple spatial scales. SADs from the SENM were found to be parameterised by a single parameter equivalent to the ratio A/R, whereby A is the number of individuals in the sample and R is proportional to a species' spatial range (Rosindell & Cornell, 2013). As with the SINM, the SENM predicts a

logseries SAD at the largest scales. However, the predicted shape of the SAD at scales smaller than the metacommunity differs from that of the SINM.

The mode of speciation (Table A1.2) on which a model is based can also affect the shape of the SAD under neutral dynamics. Hubbell's (2001) monograph was largely based on the idea of point mutation. Except perhaps for speciation by polyploidy, point mutation is unrealistic as it implies a proportional relationship between abundance and rate of speciation (Etienne *et al.*, 2007b). However, in a comparison of neutral models where speciation rate per species was proportional to abundance and where speciation was independent of abundance using 20 tree community datasets, Etienne *et al.* (2007b) found that datasets are better fitted by the point mutation models. This issue may have since been resolved as it has been found that a neutral model incorporating protracted speciation (Table A1.2) is more ecologically realistic and still retains a good fit to SAD data.

Table A1.2 The different speciation modes that have been incorporated into neutral models.

Speciation Mode	Synopsis	References
Point mutation	Speciation whereby each individual in the metacommunity has an equal probability of producing an offspring of a new species. Produces many rare species with lifetimes unrealistically short.	Hubbell (2001)
Random fission	Speciation occurs through a population randomly dividing into two distinct species. Produces species with lifetimes unrealistically long.	Hubbell & Lake (2003) Hubbell (2005), Etienne & Haegeman (2011)
Peripheral isolate	Divergence follows the isolation of populations. Newly arisen species have abundances drawn from a normal distribution.	Hubbell & Lake (2003), Hubbell (2005)
Generalised speciation	A generalised neutral community model incorporating numerous modes.	Haegeman & Etienne (2009)

Protracted speciation	Speciation is a gradual, drawn out process. Results in a new predicted metacommunity SAD (termed the 'difference logseries').	Rosindell *et al.* *(*2010)

A1.5 Criticisms of neutral models

Hubbell's (2001) publication resulted in a heated debate (e.g. Ricklefs, 2003, 2006; McGill, 2003b; Leigh, 2007; Clark, 2012). This is perhaps unsurprising, given that the basic assumptions of the theory ostensibly challenge many of the foundational concepts within ecology. The criticisms can be condensed into four main themes: a) underlying assumptions; b) fitting the ZSM and interpreting the parameters; c) changes through time; and d) attributing process to patterns. Each of these thematic criticisms will be discussed in turn.

A major issue is that species adaptations are known to be important over varying spatial scales (e.g. Condit *et al.,* 1996; Engen *et al.,* 2002), and deterministic processes of species interactions are known, at least partially, to determine species distributions. Even proponents of the neutral approach (e.g. Bell, 2001; Hubbell, 2001) agree that the theory fails at spatial scales where adaptations are important, i.e. fine spatial scales (see Jabot & Chave, 2011). The question then is not whether this assumption of neutral theory is true in all instances, as it evidently is not. Rather, the real question is given that the assumption is false should neutral models and the SADs they predict be discarded? Whilst, Chave (2004) suggests not, Ricklefs (2006) argues there is no escaping the falsified predictions involving ecological drift and changes through time (discussed below).

A specific issue relating to the SAD in the years immediately following Hubbell's (2001) monograph was the challenge of comparability of the fit of the ZSM with traditional SADs due to the complexity of the former's calculation. However, subsequent advances in regard to the ZSM have developed a sampling theory and derived its analytical form and

likelihood function (e.g. Vallade & Houchmandzadeh, 2003; Volkov *et al.*, 2003; Alonso & McKane, 2004; McKane *et al.*, 2004; Etienne, 2005; Etienne & Alonso, 2007). The sampling formula of Etienne (2005; Table A1.3) was a particularly significant development and represents the most robust method of fitting the predicted SAD of the SINM to observed data and of generating the model likelihood. A summary of developments in fitting the ZSM, including several of the tests of the ZSM involving the BCI data, is presented in Table A1.3.

Table A1.3 A summary of the developments in the history of the zero-sum multinomial distribution (ZSM) of the spatially implicit neutral model (SIM), and the SAD of the spatially explicit neutral model (SEM), and the attempts to fit both to the Barro Colorado Island 50ha tree dataset.

Study Authors	Best Model/Main Finding	Subsequent Criticisms
Hubbell (2001)	ZSM	Goodness of fit only determined by graphical observation.
McGill (2003)	ZSM does not fit the data better than the lognormal.	Used simulations to fit the ZSM.
Volkov *et al.* (2003)	Derived an analytical solution for the ZSM and found the ZSM provided the best fit.	Analytical equations did not represent the full solution as they applied solely to the mean number of species in a given class (Etienne & Olff, 2004).
Vallade & Houchmandzadeh (2003)	Published a full analytical solution for the ZSM.	Equations were later determined to be flawed (i.e. they applied the mean number of species in a given class) and were corrected by Etienne & Alonso (2005).
Alonso & McKane (2004)	Developed a different analytic solution.	Rigorous fitting of the ZSM required likelihood methods.
Etienne & Olff (2004)	Found slightly better support for the lognormal using a Bayesian approach.	

Etienne (2005)	Published the correct analytical solution and sampling formula. Two forms of the likelihood equations exist: a) Ewens' (1972) sampling formula of neutral alleles is used in the case of no dispersal limitation, and b) Etienne's (2005) formula in cases of dispersal limitation.
Etienne & Alonso (2005)	Unified two different approaches to arrive at the full analytical solution: the genealogical approach (Etienne, 2005) and master-equation based approach (e.g. Alonso & McKane, 2004).
McGill *et al.* (2006)	Compared nine goodness-of-fit measures with the BCI data and found that for eight out of the nine measures the lognormal outperformed the ZSM.
Jabot & Chave (2011)	Built on Etienne's (2005) maximum likelihood framework to develop a more robust test of neutrality incorporating the SAD of Hubbell's SINM and Shannon's index. The SAD of the BCI 50ha plot did not significantly differ from neutrality; however, the SADs of smaller scale subplots from within the BCI plot were significantly non-neutral.
Rosindell & Cornell (2013)	a) The gamma and negative binomial distributions provided a better fit than the ZSM. b) The SENM predicts SADs which are more realistic than those from the SINM.

McGill *et al.* (2006) outline two further issues with fitting the ZSM. First, as the ZSM has three parameters independent of each other, it is particularly flexible in terms of fitting, and a good fit to a particular dataset is not surprising. Second, and despite the biological meaning of the parameters being one of the original perceived benefits of neutral theory (Hubbell, 2001), the parameters can be exceedingly difficult to calculate. While calculating

the size of the local population (J) is an achievable aim, deducing the speciation rate (*v*) is likely a near impossible task (but see Ricklefs, 2006), and calculation of the metapopulation size (Jm) is contingent on being able to effectively delineate the metapopulation boundaries (Ricklefs, 2003; McGill *et al.,* 2006). However, it is important to remember that the ZSM is the SAD predicted by a particular neutral model, Hubbell's (2001) SINM. The recent finding of Rosindell & Cornell (2013) that the SADs of the SENM are more realistic than the ZSM may overcome many of these issues and warrants further study. For instance, SENMs are advantageous in regard to parameter estimation as they do not require calculation of the metapopulation size; but they do still require estimation of *v*.

Methodological problems were also found to impact the predictions of studies incorporating neutral dynamics (see Connolly & Dornelas, 2011; Matthews & Whittaker, 2014 for a discussion of methodological issues in SAD research more generally). For instance, Latimer *et al. (*2005) used an approximation of the likelihood function of Hubbell's neutral model in combination with Bayesian methods to answer questions related to immigration and speciation in the context of SADs in the Cape Floristic Region (CPF) biodiversity hotspot, South Africa. It was concluded that the neutral model can corroborate theories of high speciation rates and low migration within the region (Latimer *et al.,* 2005). However, in a reply to this paper, Etienne *et al.* (2006) argued that the exact likelihood function should be used as it enables more effective searching of the parameter space. For one of the three CPF datasets analysed by Latimer *et al.* (2005) it was found that there were in fact two similar maxima in the likelihood and that using the exact likelihood resulted in different parameter values to those of Latimer *et al.* (Etienne *et al.,* 2006). These multiple peaks in the likelihood relate to different ecological situations (i.e. different parameter values) and it is important to determine the actual maximum likelihood for a given study system (e.g. Connolly *et al.,* 2009; Matthews & Whittaker, 2014).

A problem with neutral theory not confined to SADs is the predictions the theory makes relating to changes over time, particularly population changes through time and the life span of species (Leigh, 2007). For instance, Ricklefs (2006) estimates that a European passerine species population of 14.43×10^6 individuals with a generation time of three years would take more than 86×10^6 years to become extinct under a neutral model of random drift. This is clearly too long based on the existing best knowledge of species life span estimates (Rosenzweig, 1995; Ricklefs, 2003, 2006). Thus factors other than random drift must be operating, such as changes in climate. That being said, most of this criticism has been directed to neutral models incorporating unrealistic modes of speciation. Incorporating protracted speciation into a neutral model has been found to greatly improve estimates of speciation rates and species lifetimes (Rosindell *et al.*, 2010).

Finally, and as for many other ecological theories, the presence of a particular pattern does not provide proof of a particular process (McGill *et al.*, 2007). Neutral patterns can arise from non-neutral mechanisms (Purves & Pacala, 2005; Alonso *et al.*, 2006). For example, complex ecological interactions in conjunction with variable natural conditions may prevent actual competitive differences between species being expressed, leading to non-equivalent species behaving neutrally (see Alonso *et al.*, 2006). It has also been argued that diversity patterns such as the SAD are not diagnostic ecological tools (e.g. McGill, 2003a, Leigh, 2007; Clark, 2012). The SAD, in addition to other diversity patterns such as the SAR, is an aggregated (i.e. macroecological) pattern in which the property of interest is characterised by integrating over a group of species. This aggregation has been criticised as resulting in a loss of information, which subsequently reduces the power of SADs to distinguish between communities that are influenced primarily by stochastic processes and those which are influenced by mainly deterministic (niche) processes (Chisolm & Pacala, 2010, Clark, 2012; see also discussion in Pueyo *et al.*, 2007). For instance, it has been suggested that neutral

patterns can emerge as a result of averaging over a group of species that are ecologically non-neutral and which individually may differ considerably from the average behaviour (Pueyo *et al.,* 2007; Bowler & Kelly, 2012). Thus, it is possible that the success of neutral theory in tests focused on aggregated patterns such as SADs and SARs may be explained in part by this averaging process (see Bowler & Kelly, 2012); and thus the goodness of fit of SADs predicted by neutral models cannot be used as incontrovertible evidence for the absence of niche structure (Chisholm & Pacala, 2010). That being said, SADs have been shown to provide useful information regarding community structure (e.g. Ugland & Gray, 1982; Jabot & Chave, 2011; Chust *et al.,* 2013; Matthews *et al.,* 2014), and as presence–absence and abundance data are often all that is available for many ecological datasets we would argue that focusing on the SAD may still prove enlightening.

A1.6 The other side of the coin: a focus on unifying niche and neutral perspectives

Despite attracting much criticism, it may be argued that neutral theory is beneficial in that it allows for better testing and development of predictions surrounding community structure. It provides a dynamic sampling theory of community assembly founded on key processes, namely dispersal, speciation, birth and death. Several studies have used neutral models to gain insights into the mechanisms driving diversity patterns. For instance, a neutral metacommunity model revealed the fundamental role of dispersal in generating spatial diversity patterns in the Mississippi–Missouri River System, and indicated that it was not necessary to incorporate species differences into the model to predict large scale diversity patterns in this system (Muneepeerakul *et al.,* 2008). For marine phytoplankton communities, plotting the *m* parameter of the SINM neutral model against latitude revealed that immigration probability is typically lower in tropical, relative to temperate communities (Chust *et al.,* 2013). Failure of a neutral model to fit empirical data can also be revealing. To take one example, Gilbert *et al.* (2006) used a neutral model to predict the effects of forest

fragmentation on an Amazonian tree community. It was found that the model accurately predicted the number of local extinctions, but failed to predict rates of change in species composition, highlighting that species life histories (incorporating factors such as matrix tolerance) are a vital consideration when determining the impacts of fragmentation on biodiversity (for further examples see Rosindell *et al.,* 2012).

Work on neutral theory has provided a new lens through which to view SADs and other primary macroecological patterns, and has provided the stimulus for increased debate on SADs in general (Chave, 2004), particularly the mathematical methods used to test and investigate them. This has led to an increased recognition of stochastic models in ecology as an alternative to the deterministic models that have dominated the literature. Furthermore, much of the criticism directed at the neutrality assumption is misguided, as it in fact relates to the speciation mode (i.e. point mutation) incorporated in the early SINMs (Rosindell *et al.,* 2010). Switching point mutation with the more realistic protracted speciation mode has been found to greatly improve predictions, such as species lifetimes, which did not previously match empirical data (e.g. Ricklefs, 2003, 2006), while retaining the fit to SAD data.

A look to the future raises the possibility that the incorporation of neutrality within niche theory research may be less contentious, as while many view neutral and niche based theories as incompatible, this is not the case. The predominant view that niche and neutral processes are mutually exclusive is a 'false dichotomy' (Leibold & McPeek, 2006; Adler *et al.*, 2007) and ignores the possibility that both types of processes act to structure communities concurrently; a fact acknowledged by Hubbell (2001; p.24) himself. As Alonso *et al. (*2006; p.455) contend, it is likely that the ecological reality is somewhere in between: "ecological communities are not often neutral, but they are not strictly hierarchical competitive communities either," – a position closer to weak neutrality than hard neutrality, reflecting both niche related processes and neutrality. To put it another way, community structure can

be viewed as a continuum of combinations of stabilising niche mechanisms and fitness differences between species. In this context neutral models are just a special case in which fitness of all species is equal and there are no stabilizing niche effects (Adler *et al.,* 2007). For instance, it has been shown that neutrality increases with species richness in ecological systems (Gravel *et al.,* 2006; Bar-Massada *et al.,* 2014), ostensibly due to increased niche overlap, as opposed to increased niche packing, with increasing richness. To take another example, Chust *et al.* (2013) used the method of Jabot & Chave (2011; see Table A1.3), incorporating the SAD to show that marine phytoplankton communities are structured through both niche and neutral assembly processes. In the midst of the niche vs. neutrality debate several studies have looked at ways of unifying the two perspectives (Bonsall *et al.,* 2004; Tilman, 2004; Hubbell, 2006; Leibold & McPeek, 2006; Gravel *et al.,* 2006; Scheffer & van Nes, 2006; Adler *et al.,* 2007; Chisolm & Pacala, 2010, Siepielski *et al.,* 2010, Haegeman & Loreau, 2011; Vergnon *et al.,* 2012; Bar-Massada *et al.,* 2014).

Work on reconciling neutral and niche theory has taken the form of both quantitative and non-quantitative models. In relation to the former, Chisolm & Pacala (2010) proposed a SINM which incorporates niches. Within the model all the species in the metacommunity are assigned to a particular niche; but within each niche species behave neutrally. Using this model they predicted SADs that approximate those predicted by a standard neutral model in high diversity communities. This is intriguing as it suggests neutral processes underpin the SAD even when a community has strong niche structure. However, it has since been argued that the model of Chisolm & Pacala (2010) is flawed as it considers species to be largely independent (i.e. non-interacting) and is thus basically just a form of neutral model (Haegeman & Etienne, 2011; Haegeman & Loreau, 2011).

Tilman's (2004) stochastic niche theory is an earlier example of such an approach, offering an interpretation of community structure resolving some of the debate surrounding

niche and neutral based theories. Tilman proposed a theory having a basis in niche apportionment and resource use, but combining this with stochastic processes. It incorporates the ideas of dispersal and chance, and demographic stochasticity inherent within neutral theory, but equally emphasises the importance of competition and resource division. Tilman (2004) concludes by highlighting that both his stochastic niche theory and neutral theory will produce similar SAD curve shapes and thus it is the underlying assumptions that are important: stochastic niche theory assumes a correlation between traits and abundances, and environmental context, whereas neutral theory assumes no such correlation. While a useful start point, Tilman's (2004) stochastic niche theory is not a 'complete reconciliation' between competition and neutral theories as it does not incorporate the effects of variation in immigration and dispersal limitation (Gravel *et al.,* 2006). To circumvent this issue, Gravel *et al.* (2006) proposed a model unifying Tilman's theory with the dispersal element of neutral theory, resulting in a continuum from niche to full neutrality.

It has also been shown empirically that the coexistence of ecologically equivalent species is not discordant with traditional niche based theories. In an interesting test, Siepielski *et al.* (2010) manipulated the relative species abundance of two phylogenetically distant species of *Enallagma* damselflies and found evidence of ecological equivalence. However, previous work on *Enallagma* and the sister genus *Ischnura* had found strong evidence of ecological differentiation between species in the two genera, which enabled coexistence in the same food web. Thus, taken together these findings seemingly indicate that communities can be structured through both niche and neutral based processes (Siepielski *et al.,* 2010; see also Leibold & McPeek, 2006). However, given that we are unable to measure the full niche hyperspace volume, a conclusion of ecological equivalence has to be interpreted carefully.

To take one final example, the recent advancement of emergent neutrality (EN) theory (originally termed 'self-organised similarity'), that is, a situation in which neutrality is the

outcome of community evolution, has placed a renewed focus on unifying niche and neutral perspectives in ecology. EN is based on Lotka-Volterra type competition models and describes a situation in which a community of competing species is clumped into groups of coexisting species along a niche axis (Scheffer & van Nes, 2006; Segura *et al.,* 2011; Vergnon *et al.,* 2012; see also Holt, 2006). According to the EN model, each species is located at some point along the niche axis (Fig. A1.3a), and species closer together experience a greater degree of interspecific competition. Through time the impact of this interspecific competition may lead to reduced survivorship and thus to a shift in the position of species along the niche axis away from the area of greatest competition. The counter-intuitive outcome is to create a packing effect in which species are grouped into sets of "self-organised modes that contain multiple coexisting species." (Vergnon *et al.,* 2012, p.2). Within each self-organised mode species are essentially functionally equivalent (neutral); it is demographic stochasticity and immigration into the community, rather than species interactions, which are the dominating processes acting within the modes (Barabás *et al.*, 2013). This coexistence is exhibited for thousands of generations and the theory is applicable to multiple systems as long as the community is species-rich and there is a meaningful degree of niche overlap (Scheffer & van Nes, 2006).Thus, EN theory can be seen as incorporating ecological asymmetric theories, combining both ecological and evolutionary processes, and both niche and neutral perspectives; the end-point being clusters of self-similar species (Segura *et al.*, 2011). Early tests of EN have produced promising results that should provide a catalyst for future research (Segura *et al.,* 2011; Vergnon *et al.,* 2012; but see Barabás *et al.*, 2013 for a critique of the EN model).

In relation to SADs, EN theory provides an explanation of multimodality (Vergnon *et al.*, 2012; but see Barabás *et al.*, 2013). While the possibility has long been recognised, it has become increasingly apparent that many SADs may in fact be multimodal; that is,

characterised by multiple distinct modes (Ugland & Gray, 1982; Dornelas & Connolly, 2008; Matthews *et al.,* 2014). EN predicts a bimodal SAD as the species within the 'core' of the niche axis modes are all relatively abundant (Fig. A1.3), whereas the species in the 'valleys' of the niche axis modes, termed the 'outsiders', are relatively rare (Vergnon *et al.*, 2012). EN theory can also account for multimodal SADs if abundances of species within the different niche axis modes significantly differ (Fig. A1.3a). While neutral simulations have predicted multimodal SADs (Borda-de-Água *et al.,* 2007; Barabás *et al.,* 2013), it has been argued that the patchy record of empirical tests of the ZSM (Table A1.3) means that EN provides a more 'credible' interpretation of multimodal SADs than does the ZSM (Vergnon *et al.*, 2013). That being said, a number of non-neutral mechanisms have been argued to underpin multimodal SADs (e.g. Magurran & Henderson 2003; Matthews *et al.,* 2014), and it has been argued that multimodal SADs do not represent unequivocal evidence for emergent neutrality theory (Barabás *et al.,* 2013).

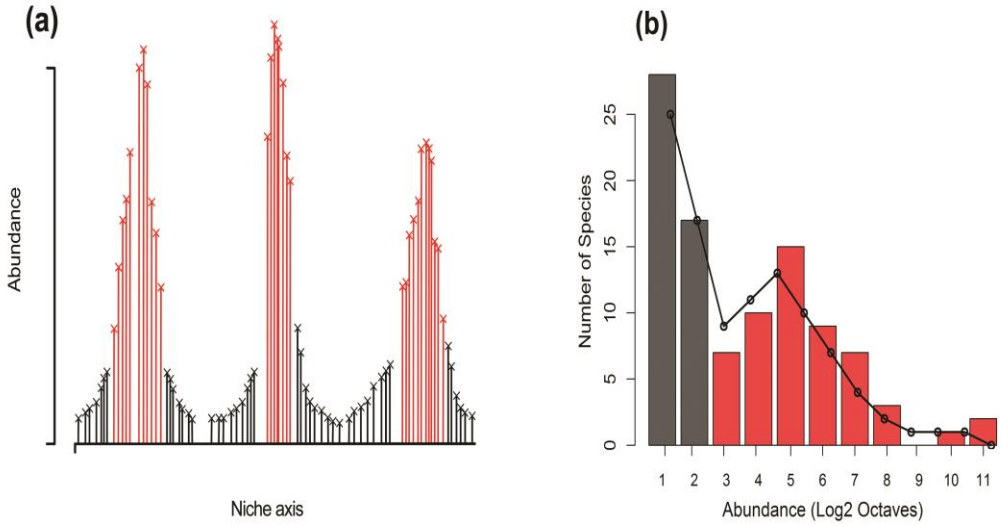


Figure A1.3 An illustration of how emergent neutrality can lead to multimodal species abundance distributions. (a) represents the abundance of a set of species as a function of a hypothetical niche axis. The species which comprise the peaks (red bars) within panel (a) are the abundant species in the community and correspond to the abundant species in the multimodal distribution (b; red bars). The species in the troughs (black bars) of panel (a) are relatively rare and correspond to the black bars in panel (b). The combination of these two sets of species in a sample results in a bimodal abundance distribution. The observe data used to construct (b) (coloured bars) are from a sample of arthropod species in a fragment of native Laurisilva forest in the Azores (P.A.V. Borges, personal communication). A two mode Poisson lognormal distribution has been fitted to the data (black line) using the functions in Dornelas & Connolly (2008). These data are used simply to provide an example of a multimodal species-abundance distribution; the role of emergent neutrality in this particular system is unclear.

A1.7 Future research avenues

Analysis of the SAD is a good test of neutral theory (Alonso *et al.*, 2006). However, early use of the ZSM distribution predicted by SINMs was hampered by the lack of a full analytical solution. The publication of the sampling formula and maximum likelihood function (e.g. Etienne, 2005) means it is now possible to fit the ZSM and compare it with other distributions (but see discussion below). The development of more spatially realistic SENMs (e.g. Rosindell & Cornell, 2013) presents additional predicted SADs, which can be used in

this endeavour. SENMs are an exciting advancement for neutral theory and SAD research more generally for two main reasons. First, they provide a form of neutral theory that predicts more realistic SADs. Second, it removes the two-tier community structure underpinning the SINM (Fig. A1.1) and thus allows exploration of the SAD at multiple scales. The spatial scaling of SADs is an exciting area of current SAD research (e.g. Borda-de-Água *et al.*, 2012) and it should be informative to gain insights from a neutral perspective in this on-going discussion.

Prior to the development of SENMs, work on neutral theory and the SAD was stifled by the niche vs. neutrality debate. In order to move forward we need to set out a new research agenda synthesising the two perspectives. Niche and neutral theories are not mutually exclusive; rather, they consider complementary processes. As such, increased insight into one theory aids in the comprehension of the other (Adler *et al.*, 2007).

Within this new research agenda it will be necessary to move away from the polarising debate in which niche models are 'pitted' against neutral models and the fit of predicted diversity patterns compared. Such an approach is based on the false dichotomisation of the two theories as wholly conflicting explanations of community dynamics. This has been shown to be an inaccurate portrayal of the ecological reality in certain cases (e.g. Siepelski *et al.*, 2010), illustrating the need to move to a layered approach in which in the roles of both niche and neutral processes are considered in tandem (e.g. Gravel *et al.*, 2006; Vergnon *et al.*, 2012). It also requires a shift away from the current approach in which evidence of neutrality is used as a beating stick with which to argue against any role of niche structure, and *vice versa* (Siepelski *et al.*, 2010). Rather, we need to work towards better integrating neutral models with classical niche coexistence models to understand the processes underpinning SADs, in addition to other macroecological patterns. Such an approach will also have applied benefits. For example, niche and neutral theories

have different implications for biodiversity conservation and a unified framework may aid in translating ecological theory to conservation practice (Holt, 2006).

A focus on integrated models will require careful consideration of the speciation mode as this is integral in determining the ecological equivalence of species within a particular taxon (Leibold & McPeek, 2006). Integrated models will also need to vary the assumptions of classical niche and neutral theories. It is likely that the ecological reality involves communities with an underlying niche structure, but with functional groups within the community that contain several ecologically similar species (Siepielski *et al.*, 2010). For example, as Haegeman & Etienne (2011, p.962) note, answering the question of how niche and neutral processes interact requires models which relax the condition of species independence (i.e. non-interaction) but in which species "interact differently when they belong to the same niche than when they belong to different niches."

In pursuit of this aim of integration it will also be necessary to consider approaches in addition to those involving the SAD. This is particularly pertinent for the SINM as analysis of the likelihood surface of the parameters m and θ has revealed a ridge where the likeliness is relatively homogenous (Etienne & Alonso, 2007). We want to determine the relative importance of niche and neutral processes in structuring communities, i.e. at which point along the aforementioned continuum of community structure particular communities lie (Gravel *et al.*, 2006; Adler *et al.*, 2007; Bar-Massada *et al.*, 2014). As SADs are often unable to discriminate between different underlying processes (above), the comparison of SAD models may have limited power in certain cases (Leibold & McPeek, 2006). Ultimately, we will require large scale manipulative experiments to provide the answers to some of these questions (Adler *et al.*, 2007). Nonetheless, SADs still have an important role to play. It is simply that to move forward we need to use abundance data to answer questions other than whether the ZSM fits the BCI data better than competing SAD models. For instance, the

prediction of multimodal SADs (Matthews *et al.*, 2014) is one of the strengths of EN theory (above).

Aside from unifying niche and neutral perspectives, neutral models and the SADs they predict can be used to gain new insights into, and make new predictions relating to, numerous classical ecological theories. For instance, a recently published neutral model of island biogeography (Rosindell & Phillimore, 2011; see also Rosindell & Harmon, 2013) builds on MacArthur & Wilson's (1967) ETIB. It produced many of the same predictions as the ETIB, such as a decrease in immigration with increasing island isolation, but also made several new predictions. For example, whereas abundance was not explicitly incorporated in MacArthur & Wilson's (1967) original model, the neutral model predicts the SAD of species on islands of varying isolation from the mainland. Finally, while it has not been extensively discussed in this review, one of the assets of neutral theory is that it predicts many ecological and biogeographical patterns (e.g. SARs; Rosindell & Cornell, 2007), not just SADs. Thus, it is important that future research, particularly the development of novel neutral models, places increased focus on the prediction of these other patterns. It is equally important that these developments incorporate recent advancements that have increased the realism of neutral theory, such as protracted speciation.

A1.8 Conclusions

In the twelve years since Hubbell (2001) published his monograph, great strides have been made in developing a robust and analytical neutral theory for ecology. In relation to SADs this has seen the derivation of an analytical solution for the ZSM with an associated sampling theory (Etienne, 2005), the use of novel approaches to determine the goodness of fit of SAD models (e.g. Etienne & Olff, 2004) and the development of SENMs (e.g. Rosindell & Cornell, 2013). Wider application of neutral theory has also been witnessed, along with

application of the ZSM beyond the theoretical domain to actual ecological problems (e.g. Latimer *et al.,* 2005); although this has not been without controversy. Due to the underlying assumptions, neutral theory and especially the ZSM have received heavy criticism (e.g. Ricklefs, 2003). However, falsifying these assumptions does not render the theory invalid in all contexts. Neutral theory in its entirety may be a simplification of ecological reality and its assumptions may often be violated in real natural systems, but its utility lies in its simplicity. Out of this controversy have arisen new perspectives on neutral theory and SADs, with a greater focus on stochastic processes. Several authors (e.g. Tilman, 2004; Gravel *et al.*, 2006; Vergnon *et al.*, 2012) have worked to develop new theories integrating aspects of both neutral and niche based theories. These theories have then been used to make predictions and explain patterns in SADs. Advancement of models acting to unify niche and neutral perspectives in ecology represents a promising avenue for future research in neutral theory and the SAD, and ultimately will provide new insights in the search for a mechanistic explanation of SADs.

A1.9 References

Adler, P.B., HilleRisLambers, J. & Levine, J.M. (2007) A niche for neutrality. *Ecology Letters*, **10**, 95-104.

Alonso, D. & McKane, A.J. (2004) Sampling Hubbell's neutral theory of biodiversity. *Ecology Letters*, **7,** 901-910.

Alonso, D., Etienne, R.S. & McKane, A.J. (2006) The merits of neutral theory. *Trends in Ecology & Evolution*, **21**, 451-457.

Alonso, D., Ostling, A. & Etienne, R.S. (2008) The implicit assumption of symmetry and the species abundance distribution. *Ecology Letters*, **11**, 93-105.

Bar-Massada, A., Kent, R. & Carmel, Y. (2014) Environmental heterogeneity affects the location of modelled communities along the niche–neutrality continuum. *Proceedings of the Royal Society B: Biological Sciences*, **281**, 20133249.

Barabás, G., D'Andrea, R., Rael, R., Meszéna, G. & Ostling, A. (2013) Emergent neutrality or hidden niches? *Oikos*, **122**, 1565-1572.

Begon, M., Harper, J.L. & Townsend, C.R. (1990) *Ecology: Individuals, populations and communities*, Second edn. Blackwell Scientific Publications, Boston, USA.

Bell, G. (2001) Neutral macroecology. *Science*, **293**, 2413-2418.

Bonsall, M.B., Jansen, V.A.A. & Hassell, M.P. (2004) Life history trade-offs assemble ecological guilds. *Science*, **306**, 111-114.

Borda-de-Água, L., Hubbell, S.P. & He, F. (2007) Scaling biodiversity under neutrality. *Scaling biodiversity* (ed. by D. Storch, P.A. Marquet and J.H. Brown), pp. 347-375. Cambridge University Press, Cambridge.

Borda-de-Água, L., Borges, P.A.V., Hubbell, S.P. & Pereira, H.M. (2012) Spatial scaling of species abundance distributions. *Ecography*, **35**, 549-556.

Bowler, M.G. & Kelly, C.K. (2012) On the statistical mechanics of species abundance distributions. *Theoretical Population Biology*, **82**, 85-91.

Caswell, H. (1976) Community structure: a neutral model analysis. *Ecological Monographs*, **46**, 327-354.

Chave, J. (2004) Neutral theory and community ecology. *Ecology Letters*, **7**, 241-253.

Chave, J. & Leigh, E.G. Jr (2002) A spatially explicit neutral model of β-diversity in tropical forests. *Theoretical Population Biology*, **62**, 153-168.

Chesson, P. (1991) A need for niches? *Trends in Ecology & Evolution*, **6**, 26-28.

Chisholm, R.A. & Pacala, S.W. (2010) Niche and neutral models predict asymptotically equivalent species abundance distributions in high-diversity ecological communities. *Proceedings of the National Academy of Sciences USA*, **107**, 15821-15825.

Chust, G., Irigoien, X., Chave, J. & Harris, R.P. (2013) Latitudinal phytoplankton distribution and the neutral theory of biodiversity. *Global Ecology & Biogeography*, **22**, 531-543.

Clark, J.S. (2012) The coherence problem with the Unified Neutral Theory of Biodiversity. *Trends in Ecology & Evolution*, **27**, 198-202.

Condit, R., Hubbell, S.P. & Foster, R.B. (1996) Changes in tree species abundance in a Neotropical forest: impact of climate change. *Journal of Tropical Ecology*, **12**, 231-256.

Connolly, S.R. & Dornelas, M. (2011) Fitting and empirical evaluation of models for species abundance distributions. In: *Biological diversity: frontiers in measurement and assessment* (ed. by A.E. Magurran and B.J. McGill), pp. 123-140. Oxford University Press, Oxford.

Connolly, S.R., Dornelas, M., Bellwood, D.R. & Hughes, T.P. (2009) Testing species abundance models: a new bootstrap approach applied to Indo-Pacific coral reefs. *Ecology*, **90**, 3138-3149.

Dornelas, M. & Connolly, S.R. (2008) Multiple modes in a coral species abundance distribution. *Ecology Letters*, **11**, 1008-1016.

Engen, S., Lande, R., Walla, T. & DeVries, P.J. (2002) Analyzing spatial structure of communities using the two-dimensional Poisson lognormal species abundance model. *American Naturalist*, **160**, 60-73.

Etienne, R.S. (2005) A new sampling formula for neutral biodiversity. *Ecology Letters*, **8**, 253-260.

Etienne, R.S. & Olff, H. (2004) A novel genealogical approach to neutral biodiversity theory. *Ecology Letters*, **7**, 170-175.

Etienne, R.S. & Alonso, D. (2005) A dispersal-limited sampling theory for species and alleles. *Ecology Letters*, **8,** 1147-1156.

Etienne, R.S. & Alonso, D. (2007) Neutral community theory: how stochasticity and dispersal-limitation can explain species coexistence. *Journal of Statistical Physics*, **128**, 485-510.

Etienne, R.S. & Haegeman, B. (2011) The neutral theory of biodiversity with random fission speciation. *Theoretical Ecology*, **4**, 87-109.

Etienne, R.S., Latimer, A.M., Silander, J.A. & Cowling, R.M. (2006) Comment on "neutral ecological theory reveals isolation and rapid speciation in a biodiversity hot spot". *Science*, **311**, 610.

Etienne, R.S., Alonso, D. & McKane, A.J. (2007a) The zero-sum assumption in neutral biodiversity theory. *Journal of Theoretical Biology*, **248**, 522-536.

Etienne, R.S., Apol, M.E.F., Olff, H. & Weissing, F.J. (2007b) Modes of speciation and the neutral theory of biodiversity. *Oikos*, **116,** 241-258.

Ewens, W.J. (1972) The sampling theory of selectively neutral alleles. *Theoretical Population Biology*, **3**, 87-112.

Fisher, R.A., Corbet, A.S. & Williams, C.B. (1943) The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology*, **12**, 42-58.

Gilbert, B., Laurance, W.F., Leigh, E.G. Jr & Nascimento, H.E.M. (2006) Can neutral theory predict the responses of Amazonian tree communities to forest fragmentation? *The American Naturalist*, **168**, 304-317.

Gotelli, N.J. & McGill, B.J. (2006) Null versus neutral models: what's the difference? *Ecography*, **29**, 793-800.

Gravel, D., Canham, C.D., Beaudet, M. & Messier, C. (2006) Reconciling niche and neutrality: the continuum hypothesis. *Ecology Letters*, **9**, 399-409.

Gray, J.S., Bjørgesæter, A. & Ugland, K.I. (2006a) On plotting species abundance distributions. *Journal of Animal Ecology*, **75**, 752-756.

Haegeman, B. & Etienne, R. (2009) Neutral models with generalised speciation. *Bulletin of Mathematical Biology*, **71**, 1507-1519.

Haegeman, B. & Etienne, R.S. (2011) Independent species in independent niches behave neutrally. *Oikos*, **120**, 961-963.

Haegeman, B. & Loreau, M. (2011) A mathematical synthesis of niche and neutral theories in community ecology. *Journal of Theoretical Biology*, **269**, 150-165.

He, F. & Hu, X.-S. (2005) Hubbell's fundamental biodiversity parameter and the Simpson diversity index. *Ecology Letters*, **8**, 386-390.

Holt, R.D. (2006) Emergent neutrality. *Trends in Ecology & Evolution*, **21**, 531-533.

Hubbell, S.P. (2001) *The unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton.

Hubbell, S.P. (2005) Neutral theory in community ecology and the hypothesis of functional equivalence. *Functional Ecology*, **19**, 166-172.

Hubbell, S.P. (2006) Neutral theory and the evolution of ecological equivalence. *Ecology*, **87,** 1387-1398.

Hubbell, S.P. & Foster, R.B. (1986) Commonness and rarity in a Neotropical forest: implications for tropical tree conservation. In: *Conservation biology* (ed. by M.E. Soulé), pp. 205-231. Sinauer Associates, Sunderland.

Hubbell, S.P. & Lake, J. (2003) The neutral theory of biodiversity and biogeography, and beyond. In: *Macroecology: concepts and consequences* (eds. T.M. Blackburn and K.J. Gaston), pp. 45-63. Cambridge University Press, Cambridge.

Hubbell, S.P., He, F., Condit, R., Borda-de-Água, L., Kellner, J. & ter Steege, H. (2008) How many tree species are there in the Amazon and how many of them will go extinct? *Proceedings of the National Academy of Sciences USA*, **105**, 11498-11504.

Jabot, F. & Chave, J. (2011) Analyzing tropical forest tree species abundance distributions using a nonneutral model and through approximate Bayesian inference. *American Naturalist*, **178**, E37-E47.

Kimura, M. (1968) Evolutionary rate at the molecular level. *Nature*, **217**, 624-626.

Kimura, M. & Crow, J.F. (1964) The number of alleles that can be maintained in a finite population. *Genetics*, **49**, 725-738.

King, J.L. & Jukes, T.H. (1969) Non-Darwinian evolution. *Science*, **164**, 788-798.

Latimer, A.M., Silander, J.A. Jr & Cowling, R.M. (2005) Neutral ecological theory reveals isolation and rapid speciation in a biodiversity hot spot. *Science*, **309**, 1722-1725.

Leibold, M.A. & McPeek, M.A. (2006) Coexistence of the niche and neutral perspectives in community ecology. *Ecology*, **87**, 1399-1410.

Leigh, E.G. Jr (2007) Neutral theory: a historical perspective. *Journal of Evolutionary Biology*, **20**, 2075-2091.

MacArthur, R.H. & Wilson, E.O. (1967) *The theory of island biogeography*. Princeton University Press, Princeton.

Magurran, A.E. (2005) Species abundance distributions: pattern or process? *Functional Ecology*, **19,** 177-181.

Magurran, A.E. & Henderson, P.A. (2003) Explaining the excess of rare species in natural species abundance distributions. *Nature*, **422**, 714-716.

Matthews, T.J. & Whittaker, R.J. (2014) Species abundance distributions: a review of recent advances in approaches for fitting and comparing competing models. *Frontiers of Biogeography*, **6**, 67-82.

Matthews, T.J., Borges, P.A.V. & Whittaker R.J. (2014) Multimodal species abundance distributions: a deconstruction approach reveals the processes behind the pattern. *Oikos,* **123,** 533-544.

McGill, B. (2003a) Strong and weak tests of macroecological theory. *Oikos,* **102,** 679-685.

McGill, B.J. (2003b) A test of the unified neutral theory of biodiversity. *Nature,* **422,** 881-885.

McGill, B.J. (2006) A renaissance in the study of abundance. *Science,* **314**, 770-772.

McGill, B.J. (2011) Species abundance distributions. *Biological diversity: frontiers in measurement and assessment* (ed. by A.E. Magurran and B.J. McGill), pp. 105-122. Oxford University Press, Oxford.

McGill, B.J., Maurer, B.A. & Weiser, M.D. (2006) Empirical evaluation of neutral theory. *Ecology*, **87**, 1411-1423.

McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K., Dornelas, M., Enquist, B.J., Green, J.L., He, F., Hurlbert, A.H., Magurran, A.E., Marquet, P.A., Maurer, B.A., Ostling, A., Soykan, C.U., Ugland, K.I. & White, E.P. (2007) Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters*, **10**, 995-1015.

McKane, A.J., Alonso, D. & Solé, R.V. (2004) Analytic solution of Hubbell's model of local community dynamics. *Theoretical Population Biology,* **65**, 67-73.

Muneepeerakul, R., Bertuzzo, E., Lynch, H.J., Fagan, W.F., Rinaldo, A. & Rodriguez-Iturbe, I. (2008) Neutral metacommunity models predict fish diversity patterns in Mississippi-Missouri basin. *Nature,* **453**, 220-222.

Preston, F.W. (1948) The commonness, and rarity, of species. *Ecology,* **29**, 254-283.

Pueyo, S., He, F. & Zillio, T. (2007) The maximum entropy formalism and the idiosyncratic theory of biodiversity. *Ecology Letters,* **10**, 1017-1028.

Purves, D.W. & Pacala, S.W. (2005) Ecological drift in niche-structured communities: neutral pattern does not imply neutral process. *Biotic interactions in the tropics* (ed. by D. Burslem, M. Pinard and S. Hartley), pp. 107-138. Cambridge University Press, Cambridge.

Ricklefs, R.E. (2003) A comment on Hubbell's zero-sum ecological drift model. *Oikos*, **100**, 185-192.

Ricklefs, R.E. (2006) The unified neutral theory of biodiversity: do the numbers add up? *Ecology*, **87**, 1424-1431.

Rosenzweig, M.L. (1995) *Species diversity in space and time.* Cambridge University Press, Cambridge.

Rosindell, J. & Cornell, S.J. (2007) Species–area relationships from a spatially explicit neutral model in an infinite landscape. *Ecology Letters*, **10**, 586-595.

Rosindell, J. & Cornell, S.J. (2013) Universal scaling of species-abundance distributions across multiple scales. *Oikos*, **122**, 1101-1111.

Rosindell, J. & Harmon, L.J. (2013) A unified model of species immigration, extinction and abundance on islands. *Journal of Biogeography*, **40**, 1107-1118.

Rosindell, J. & Phillimore, A.B. (2011) A unified model of island biogeography sheds light on the zone of radiation. *Ecology Letters*, **14**, 552-560.

Rosindell, J., Wong, Y. & Etienne, R.S. (2008) A coalescence approach to spatial neutral ecology. *Ecological Informatics*, **3**, 259-271.

Rosindell, J., Cornell, S.J., Hubbell, S.P. & Etienne, R.S. (2010) Protracted speciation revitalizes the neutral theory of biodiversity. *Ecology Letters*, **13**, 716-727.

Rosindell, J., Hubbell, S.P. & Etienne, R.S. (2011) The unified neutral theory of biodiversity and biogeography at age ten. *Trends in Ecology & Evolution*, **26,** 340-348.

Rosindell, J., Hubbell, S.P., He, F., Harmon, L.J. & Etienne, R.S. (2012) The case for ecological neutral theory. *Trends in Ecology & Evolution*, **27**, 203-208.

Scheffer, M. & van Nes, E.H. (2006) Self-organized similarity, the evolutionary emergence of groups of similar species. *Proceedings of the National Academy of Sciences USA*, **103**, 6230-6235.

Segura, A.M., Calliari, D., Kruk, C., Conde, D., Bonilla, S. & Fort, H. (2011) Emergent neutrality drives phytoplankton species coexistence. *Proceedings of the Royal Society B: Biological Sciences*, **278,** 2355-2361.

Siepielski, A.M., Hung, K.-L., Bein, E.E.B. & McPeek, M.A. (2010) Experimental evidence for neutral community dynamics governing an insect assemblage. *Ecology*, **91**, 847-857.

Sugihara, G. (1980) Minimal community structure: an explanation of species abundance patterns. *American Naturalist*, **116,** 770-787.

Tilman, D. (2004) Niche tradeoffs, neutrality, and community structure: a stochastic theory of resource competition, invasion, and community assembly. *Proceedings of the National Academy of Sciences USA*, **101**, 10854-10861.

Ugland, K.I. & Gray, J.S. (1982) Lognormal distributions and the concept of community equilibrium. *Oikos*, **39**, 171-178.

Ulrich, W., Ollik, M. & Ugland, K.I. (2010) A meta-analysis of species–abundance distributions. *Oikos*, **119,** 1149-1155.

Vallade, M. & Houchmandzadeh, B. (2003) Analytical solution of a neutral model of biodiversity. *Physical Review E*, **68**, 0619021 - 0619025.

Vergnon, R., van Nes, E.H. & Scheffer, M. (2012) Emergent neutrality leads to multimodal species abundance distributions. *Nature Communications*, 3, 663.

Vergnon, R., van Nes, E.H. & Scheffer, M. (2013) Interpretation and predictions of the Emergent neutrality model: a reply to Barabás *et al.*, *Oikos*, **122**, 1573-1575.

Volkov, I., Banavar, J.R., Hubbell, S.P. & Maritan, A. (2003) Neutral theory and relative species abundance in ecology. *Nature*, **424**, 1035-1037.

Watterson, G.A. (1974a) The sampling theory of selectively neutral alleles. *Advances in Applied Probability,* **6**, 463-488.

Watterson, G.A. (1974b) Models for the logarithmic species abundance distribution. *Theoretical Population Biology*, **6**, 217-250.

Whittaker, R.J. & Fernández-Palacios, J.M. (2007) *Island biogeography: ecology, evolution, and conservation.* 2nd edn. Oxford University Press, Oxford.

Appendix 2

Supporting information: Thresholds and the species–area relationship: a synthetic analysis of habitat island datasets (Chapter 4)

A2.1 Source paper information

Description of the source papers, including taxa, habitat island type (IT), species richness range (RR), island area range (AR), number of islands (NI), and the corresponding best island species–area models across each data transformation. The full references follow the table.

Study	Taxon	IT	RR	AR	NI	Untransformed	Semi-Log Transformation	Log-Log Transformation
Arroyo-Rodríguez & Mandujano (2006)	Plants	Forest	53-83	0.01-0.756	15	None Significant	None Significant	None Significant
Báldi & Kisbenedek (1999)	Invertebrates	Non-forest	1-15	0.00018-0.4	27	None Significant	Piecewise Continuous (Shallow-Steep)	Piecewise no-slope
Bauer (1989)	Invertebrates	Non-forest	46-58	0.002-0.14	8	None Significant	None Significant	None Significant
Baz *et al.* (1995)	Invertebrates	Forest	26-43	0.036-21.15	13	None Significant	None Significant	None Significant
Bell & Donelly (2006)	Reptiles	Forest	10-24	0.0143-0.0688	9	Linear Regression (Positive)	Piecewise no-slope	Piecewise no-slope
Bennet (1997)	Plants	Non-forest	174-1353	0.0324-6568	18	Piecewise Continuous (Steep-Shallow)	Piecewise Continuous (Shallow-Steep)	Piecewise Discontinuous
Blake (1986)	Birds	Forest	39-84	0.018-6	14	None Significant	None Significant	None Significant
Blake & Karr (1984)	Birds	Forest	4-24	0.018-6	12	Piecewise Continuous (Steep-Shallow)	Linear Regression (Positive)	Piecewise Discontinuous
Bowers & McLaughlin (1982)	Plants	Non-fore	19 5-15	1.26-4930	20	None Significant	Linear Regression (Positive)	Linear Regression (Positive)

		st	74					
Brown & Hutchings (1997)	Invertebrates	Forest	12 0-242	0.01-1	12	Piecewise Continuous (Steep-Shallow)	Linear Regression (Positive)	Linear Regression (Positive)
Carter-Lovejoy (1982)	Mammals	Non-forest	3-7	0.0016-0.036	14	Linear Regression (Positive)	Piecewise no-slope	Piecewise no-slope
Carter-Lovejoy (1982)	Plants	Non-forest	25-49	0.0016-0.036	14	None Significant	None Significant	None Significant
Castelletta *et al.* (2005)	Birds	Non-forest	49-98	0.07-9.35	10	Piecewise Continuous (Steep-Shallow)	Piecewise Continuous (Steep-Shallow)	Piecewise Continuous (Steep-Shallow)
Cayuela *et al.* (2006)	Plants	Forest	12-79	0.71-56.64	16	None Significant	None Significant	None Significant
Crooks (2002)	Mammals	Non-forest	2-7	0.02-1.02	29	None Significant	None Significant	None Significant
Crooks *et al.* (2001)	Birds	Non-forest	0-7	0.02-1.02	30	Linear Regression (Positive)	Linear Regression (Positive)	Piecewise Continuous (Steep-Shallow)
Crowe (1979)	Plants	Non-forest	9-69	0.000111-0.007371	26	Piecewise Continuous (Steep-Shallow)	Piecewise Discontinuous	Piecewise Discontinuous
Darlington *et al.* (2001)	Invertebrates	Forest	0-11	0.000004-0.0042	25	Piecewise Discontinuous	Linear Regression (Positive)	Linear Regression (Positive)
Dickman (1987)	Mammals	Non-forest	2-17	0.0016-0.2	50	None Significant	Piecewise Continuous (Steep-Shallow)	Piecewise Continuous (Steep-Shallow)
dos Anjos & Boçon (1999)	Birds	Forest	45-139	0.005-8.4	12	None Significant	Linear Regression (Positive)	None Significant
dos Santos *et al.* (2007)	Plants	Forest	47-110	0.124-0.634	10	None Significant	None Significant	None Significant
Dunn & Loehle (1988)	Plants	Forest	20-61	0.04-0.3	15	None Significant	None Significant	None Significant
Ferraz *et al.* (2003)	Birds	Forest	73-11	0.011-	11	Linear Regression	Linear Regression	Linear Regression

			5	1.012		(Positive)	(Positive)	(Positive)
Flaspohler *et al.* (2010)	Birds	For est	2-10	0.0007-0.557	18	Piecewise Continuous (Steep-Shallow)	Piecewise Continuous (Shallow-Steep)	Linear Regression (Positive)
Ford (1987)	Birds	For est	13-35	0.0014-0.18	20	Piecewise no-slope	Linear Regression (Positive)	Piecewise Discontinuous
Ford (1987)	Plants	Non-forest	7-17	0.0014-0.18	20	None Significant	Piecewise no-slope	Linear Regression (Positive)
Forman *et al.* (1976)	Birds	For est	0-18	0.0001-0.24	10	None Significant	Piecewise no-slope: FS	Linear Regression (Positive)
Frank & Battisti (2005)	Birds	For est	2-25	0.0012-5.2564	16	Linear Regression (Positive)	Linear Regression (Positive)	Piecewise Continuous (Steep-Shallow)
Gaublomme *et al.* (2008)	Invertebrates	Non-forest	21-56	0.053-0.431	9	None Significant	None Significant	None Significant
Godefroid & Koedam (2003)	Plants	For est	25-56	0.02-1.23	11	Piecewise Continuous (Steep-Shallow)	None Significant	Piecewise Continuous (Steep-Shallow)
Granados *et al.* (2001)	Plants	Non-forest	160-592	0.202-14.164	10	Linear Regression (Positive)	Linear Regression (Positive)	Linear Regression (Positive)
Hamel *et al.* (1993)	Birds	For est	11-23	0.17-3.04	32	None Significant	None Significant	None Significant
Hattori & Ishida (2000)	Plants	For est	19-110	0.00003-0.0784	38	Piecewise Discontinuous	Piecewise Discontinuous	Piecewise Continuous (Steep-Shallow)
Ishida *et al.* (1998)	Plants	For est	18-84	0.00025-0.164	29	None Significant	Piecewise no-slope	Linear Regression (Positive)
Kelt (2000)	Mammals	For est	3-6	0.0189-1.249	14	None Significant	None Significant	None Significant
Kitchener *et al.* (1980)	Mammals	Non-forest	2-13	0.34-51.19	23	Linear Regression (Positive)	Linear Regression (Positive)	Piecewise Continuous (Steep-Shallow)
Kitchener *et al.* (1982)	Birds	Non-fore	29-107	0.34-51.19	22	None Significant	Piecewise Continuous (Steep-	Piecewise Continuous (Steep-

		st					Shallow)	Shallow)
Kitchener *et al.* (1982)	Plants	Non-forest	37-288	0.34-51.19	22	Linear Regression (Positive)	Piecewise Continuous (Shallow-Steep)	Linear Regression (Positive)
Lawesson *et al.* (1998)	Plants	Forest	28-102	0.011-4.457	62	Linear Regression (Positive)	Linear Regression (Positive)	Linear Regression (Positive)
Lawesson *et al.* (1998)	Plants	Forest	35-192	0.011-4.457	62	Piecewise Continuous (Steep-Shallow)	Linear Regression (Positive)	Linear Regression (Positive)
Lehvävirta *et al.* (2006)	Invertebrates	Non-forest	2-23	0.009-3.4	15	Linear Regression (Negative)	Piecewise no-slope	None Significant
Levenson (1981)	Plants	Forest	2-18	0.0003-0.3996	43	Piecewise Discontinuous	Piecewise Discontinuous	None Significant
Levenson (1981)	Plants	Forest	17-42	0.0003-0.3996	43	Piecewise Discontinuous	Piecewise Discontinuous	Piecewise Discontinuous
Levenson (1981)	Plants	Forest	8-35	0.0003-0.3996	43	None Significant	Piecewise Discontinuous	Piecewise Discontinuous
Lomolino & Perault (2001)	Mammals	Forest	1-8	0.0093-0.5891	20	None Significant	None Significant	None Significant
Lövei *et al.* (2006)	Invertebrates	Forest	10-29	0.41-39.95	15	None Significant	None Significant	None Significant
Lovenzetti & Battisti (2007)	Plants	Forest	9-22	0.0026-3.0248	20	None Significant	Linear Regression (Positive)	Linear Regression (Positive)
Lovenzetti & Battisti (2007)	Birds	Forest	4-11	0.0026-3.0248	20	Piecewise Continuous (Steep-Shallow)	Linear Regression (Positive)	Piecewise Continuous (Steep-Shallow)
Loyn (1997)	Birds	Forest	8-67	0.001-97.33	56	None Significant	Piecewise Continuous (Steep-Shallow)	Piecewise Continuous (Steep-Shallow)
Lyman (1997)	Mammals	Forest	1-10	0.007-1.094	12	None Significant	Piecewise no-slope	Linear Regression (Positive)
Magura *et al.*	Invertebr	For	10-	0.411	15	None	None	None

(2001)	ates	est	29	9-39.9546		Significant	Significant	Significant
Matthiae & Strearns (1981)	Mammals	Forest	4-13	0.004-0.4	22	Piecewise Continuous (Steep-Shallow)	Piecewise no-slope	Piecewise no-slope
McFrederick & LeBuhn (2006)	Invertebrates	Forest	1-3	0.0032-1.5864	18	None Significant	None Significant	None Significant
Metzger *et al.* (1997)	Plants	Forest	14-111	0.0063-0.7353	15	None Significant	Linear Regression (Positive)	Linear Regression (Positive)
Miller & Harris (1977)	Mammals	Non-forest	29-55	114-21358	13	None Significant	None Significant	None Significant
Miyashita *et al.* (1998)	Invertebrates	Forest	7-39	0.002-50	17	None Significant	Piecewise Continuous (Steep-Shallow)	Piecewise Continuous (Steep-Shallow)
Natuhara & Imai (1999)	Birds	Forest	3-44	0.0006-647	31	None Significant	Piecewise Continuous (Shallow-Steep)	Piecewise Discontinuous
Newmark (1991)	Birds	Forest	2-18	0.001-0.3002	9	Piecewise Continuous (Steep-Shallow)	Piecewise no-slope	Linear Regression (Positive)
Onderdonk & Chapman (2000)	Mammals	Forest	1-4	0.008-1.3	20	None Significant	None Significant	None Significant
Pahl *et al.* (1988)	Mammals	Forest	1-5	0.0236-0.7449	11	Piecewise Continuous (Steep-Shallow)	Linear Regression (Positive)	Linear Regression (Positive)
Pearce *et al.* (2005)	Invertebrates	Forest	10-31	0.005-0.112	10	Piecewise Continuous (Steep-Shallow)	None Significant	None Significant
Ramanamanjato (2000)	Reptiles	Forest	6-51	0.1-4.57	11	Linear Regression (Positive)	Linear Regression (Positive)	None Significant
Recher *et al.* (1987)	Plants	Forest	2-7	0.0168-0.1273	22	Linear Regression (Positive)	Linear Regression (Positive)	Linear Regression (Positive)
Rosenblatt *et*	Mammal	For	8-	0.018	10	None	Linear	Linear

al. (1999)	s	est	14	-6		Significant	Regression (Positive)	Regression (Positive)
Shreeve & Mason (1980)	Invertebrates	Forest	1-22	0.02-1.75	22	Piecewise Continuous (Steep-Shallow)	Linear Regression (Positive)	Linear Regression (Positive)
Soulé *et al.* (1979)	Mammals	Non-forest	11-70	0.2-20808	19	None Significant	Linear Regression (Positive)	Piecewise Continuous (Steep-Shallow)
Soulé *et al.* (1988)	Birds	Non-forest	0-7	0.004-1.0277	37	Piecewise Continuous (Steep-Shallow)	Linear Regression (Positive)	Linear Regression (Positive)
Summerville *et al.* (2002)	Invertebrates	Non-forest	1-18	0.000045-0.000225	25	Linear Regression (Positive)	Linear Regression (Positive)	Linear Regression (Positive)
Tellería & Santos (1995)	Birds	Forest	1-9	0.001-3.5	27	Piecewise Continuous (Steep-Shallow)	Linear Regression (Positive)	Linear Regression (Positive)
Viveiros de Castro (2004)	Mammals	Forest	4-11	0.012-0.133	8	None Significant	None Significant	None Significant
Watson (2003)	Birds	Forest	0-56	0.02-1592.46	17	Piecewise Continuous (Steep-Shallow)	Piecewise no-slope	None Significant
Watson *et al.* (2000)	Birds	Forest	8-23	0.025-4.57	27	None Significant	None Significant	None Significant
Weaver & Kellman (1981)	Plants	Forest	8-16	0.0082-0.0811	10	None Significant	None Significant	None Significant
Western & Ssemakula (1981)	Mammals	Non-forest	15-31	54-20808	19	None Significant	None Significant	None Significant
Whitcomb *et al.* (1981)	Birds	Forest	17-28	0.011-2.83	25	Linear Regression (Positive)	Linear Regression (Positive)	Linear Regression (Positive)
Woolhouse (1983)	Birds	Forest	14-37	0.036-0.401	30	Linear Regression (Positive)	Piecewise no-slope	Piecewise no-slope

Source papers:

Arroyo-Rodríguez, V. & Mandujano, S. (2006) The importance of tropical rain forest fragments to the conservation of plant species diversity in Los Tuxtlas, Mexico. *Biodiversity & Conservation*, **15**, 4159-4179.

Báldi, A. & Kisbenedek, T. (1999) Orthopterans in small steppe patches: an investigation for the best-fit model of the species–area curve and evidences for their non-random distribution in the patches. *Acta Oecologica-International Journal of Ecology*, **20**, 125-132.

Bauer, L.J. (1989) Moorland beetle communities on limestone `habitat islands'. I. Isolation, invasion and local species diversity in carabids and staphylinids. *Journal of Animal Ecology*, **58**, 1077-1098.

Baz, A. & Garcia-Boyero, A. (1995) The effects of forest fragmentation on butterfly communities in Central Spain. *Journal of Biogeography*, **22**, 129-140.

Bell, K.E. & Donnelly, M.A. (2006) Influence of forest fragmentation on community structure of frogs and lizards in northeastern Costa Rica. *Conservation Biology*, **20**, 1750-1760.

Bennett, J.P. (1997) Nested taxa-area curves for eastern United States floras. *Rhodora*, **99**, 241-251.

Blake, J.G. (1986) Species–area relationship of migrants in isolated woodlots in east-Central Illinois. *The Wilson Bulletin*, **98**, 291-296.

Blake, J.G. & Karr, J.R. (1984) Species composition of bird communities and the conservation benefit of large versus small forests. *Biological Conservation*, **30**, 173-187.

Bowers, J.E. & Mclaughlin, S.P. (1982) Plant species diversity in Arizona. *Madroño*, **29**, 227-233.

Brown, K.S. & Hutchings, R.W. (1997) Disturbance, fragmentation, and the dynamics of diversity in amazonian forest butterflies. *Tropical forest remnants: Ecology, management and conservation of fragmented communities* (ed. by W.F. Laurance and R.O. Bierregaard), pp. 91-110. Chicago University Press, Chicago, US.

Carter-Lovejoy, S.H. (1982) The relation between species numbers and island characteristics for habitat islands in a volcanic landscape. *Great Basin Naturalist*, **42**, 113-119.

Castelletta, M., Thiollay, J.-M. & Sodhi, N.S. (2005) The effects of extreme forest fragmentation on the bird community of Singapore Island. *Biological Conservation*, **121**, 135-155.

Cayuela, L., Golicher, D.J., Benayas, J.M.R., González-Espinosa, M. & RamÍrez-Marcial, N. (2006) Fragmentation, disturbance and tree diversity conservation in tropical montane forests. *Journal of Applied Ecology*, **43**, 1172-1181.

Crooks, K.R. (2002) Relative sensitivities of mammalian carnivores to habitat fragmentation. *Conservation Biology*, **16**, 488-502.

Crooks, K.R., Suarez, A.V., Bolger, D.T. & Soulé, M.E. (2001) Extinction and colonization of birds on habitat islands. *Conservation Biology*, **15**, 159-172.

Crowe, T.M. (1979) Lots of Weeds: Insular phytogeography of vacant urban lots. *Journal of Biogeography*, **6**, 169-181.

Darlington, J.P.E.C., Kaib, M. & Brand, R. (2001) Termites (Isoptera) in forest remnants and forest Islands in the Shimba Hills National Reserve, coastal province of Kenya. *Sociobiology*, **37**, 527-538.

Dickman, C.R. (1987) Habitat fragmentation and vertebrate species richness in an urban environment. *Journal of Applied Ecology*, **24**, 337-351.

dos Anjos, L. & Boçon, R. (1999) Bird communities in natural forest patches in southern brazil. *The Wilson Bulletin*, **111**, 397-414.

dos Santos, K., Kinoshita, L.S. & dos Santos, F.A.M. (2007) Tree species composition and similarity in semideciduous forest fragments of southeastern Brazil. *Biological Conservation*, **135**, 268-277.

Dunn, C.P. & Loehle, C. (1988) Species–area parameter estimation: testing the null model of lack of relationship. *Journal of Biogeography*, **15**, 721-728.

Ferraz, G., Russell, G.J., Stouffer, P.C., Bierregaard, R.O., Pimm, S.L. & Lovejoy, T.E. (2003) Rates of species loss from Amazonian forest fragments. *Proceedings of the National Academy of Sciences USA*, **100**, 14069-14073.

Flaspohler, D.J., Giardina, C.P., Asner, G.P., Hart, P., Price, J., Lyons, C.K.A. & Castaneda, X. (2010) Long-term effects of fragmentation and fragment properties on bird species richness in Hawaiian forests. *Biological Conservation*, **143**, 280-288.

Ford, H.A. (1987) Bird communities on habitat islands in England. *Bird Study*, **34**, 205 - 218.

Forman, R.T.T., Galli, A.E. & Leck, C.F. (1976) Forest size and avian diversity in New Jersey woodlots with some land use implications. *Oecologia*, **26**, 1-8.

Frank, B. & Battisti, C. (2005) Area effect on bird communities, guilds and species in a highly fragmented forest landscape of central Italy. *Italian Journal of Zoology*, **72**, 297-304.

Gaublomme, E., Hendrickx, F., Dhuyvetter, H. & Desender, K. (2008) The effects of forest patch size and matrix type on changes in carabid beetle assemblages in an urbanized landscape. *Biological Conservation*, **141**, 2585-2596.

Godefroid, S. & Koedam, N. (2003) How important are large vs. small forest remnants for the conservation of the woodland flora in an urban context? *Global Ecology and Biogeography*, **12**, 287-298.

Granados, M., O'Kennon, R.J. & Benz, B.F. (2001) Plant species–area relationships in ten central Texas protected natural areas. *SIDA Contributions to Botany*, **19**, 1061-1072.

Hamel, P.B., Smith, W.P. & Wahl, J.W. (1993) Wintering bird populations of fragmented forest habitat in the Central Basin, Tennessee. *Biological Conservation*, **66**, 107-115.

Hattori, T. & Ishida, H. (2000) Relationship between species diversity, species composition and forest area of fragmented lucidophyllous forests in central Miyazaki Prefecture. *Japanese Journal of Ecology (Otsu)*, **50**, 221-234.

Ishida, H., Hattori, T., Takeda, Y. & Kodate, S. (1998) Relationship between species richness or species composition and area of fragmented lucidophyllous forests in southeastern Hyogo Prefecture. *Japanese Journal of Ecology (Tokyo)*, **48**, 1-16.

Kelt, D.A. (2000) Small mammal communities in rainforest fragments in Central Southern Chile. *Biological Conservation*, **92**, 345-358.

Kitchener, D.J., Dell, J., Muir, B.G. & Palmer, M. (1982) Birds in Western Australian wheatbelt reserves—implications for conservation. *Biological Conservation*, **22**, 127-163.

Kitchener, D.J., Chapman, A., Muir, B.G. & Palmer, M. (1980) The conservation value for mammals of reserves in the western australian wheatbelt. *Biological Conservation*, **18**, 179-207.

Lawesson, J.E., de Blust, G., Grashof, C., Firbank, L., Honnay, O., Hermy, M., Hobitz, P. & Jensen, L.M. (1998) Species diversity and area–relationships in Danish beech forests. *Forest Ecology and Management*, **106**, 235-245.

Lehvävirta, S., Kotze, D., Niemelä, J., Mäntysaari, M. & O'Hara, B. (2006) Effects of fragmentation and trampling on carabid beetle assemblages in urban woodlands in Helsinki, Finland. *Urban Ecosystems*, **9**, 13-26.

Levenson, J.B. (1981) Woodlots as biogeographic islands in southern Wisconsin. *Forest island dynamics in man-dominated landscapes* (ed. by R.L. Burgess and D.M. Sharpe), pp. 13-39. Springer-Verlag, New York.

Lomolino, M.V. & Perault, D.R. (2001) Island biogeography and landscape ecology of mammals inhabiting fragmented, temperate rain forests. *Global Ecology and Biogeography*, **10**, 113-132.

Lorenzetti, E. & Battisti, C. (2007) Nature reserve selection on forest fragments in a suburban landscape (Rome, Central Italy): indications from a set of avian species. *Landscape Research*, **32**, 57-78.

Lövei, G.L., Magura, T., Tothmeresz, B. & Kodobocz, V. (2006) The influence of matrix and edges on species richness patterns of ground beetles (Coleoptera : Carabidae) in habitat islands. *Global Ecology and Biogeography*, **15**, 283-289.

Loyn, R.H. (1987) Effects of patch area and habitat on bird abundances, species numbers and tree health in fragmented Victorian forests. *Nature Conservation: The Role of Remnants of Native Vegetation* (ed. by D.A. Saunders, G.W. Arnold, A.A. Burbridge and A.J.M. Hopkins), pp 65–77. Surrey Beatty and Sons, Chipping Norton.

Lyman, A.J. (1997) Rapid decline of small mammal diversity in monsoon evergreen forest fragments in Thailand. *Tropical forest remnants: Ecology, management and conservation of fragmented communities* (ed. by W.F. Laurance and R.O. Bierregaard), pp. 222-240. University of Chicago Press, Chicago, US.

Magura, T., Kodobocz, V. & Tothmeresz, B. (2001) Effects of habitat fragmentation on carabids in forest patches. *Journal of Biogeography*, **28**, 129-138.

Matthiae, P.E. & Stearns, F. (1981) Mammals in forests islands in southeastern Wisconsin. *Forest Island Dynamics in Man-Dominated Landscapes* (ed. by R.L. Burgess and D.M. Sharpe), pp. 55–66. Springer-Verlag, New York.

McFrederick, Q.S. & LeBuhn, G. (2006) Are urban parks refuges for bumble bees *Bombus* spp. (Hymenoptera: Apidae)? *Biological Conservation*, **129**, 372-382.

Metzger, J.P., Bernacci, L.C. & Goldenberg, R. (1997) Pattern of tree species diversity in riparian forest fragments of different widths (SE Brazil). *Plant Ecology*, **133**, 135-152.

Miller, R.I. & Harris, L.D. (1977) Isolation and extirpations in wildlife reserves. *Biological Conservation*, **12**, 311-315.

Miyashita, T., Shinkai, A. & Chida, T. (1998) The effects of forest fragmentation on web spider communities in urban areas. *Biological Conservation*, **86**, 357-364.

Natuhara, Y. & Imai, C. (1999) Prediction of species richness of breeding birds by landscape-level factors of urban woods in Osaka Prefecture, Japan. *Biodiversity and Conservation*, **8**, 239-253.

Newmark, W.D. (1991) Tropical forest fragmentation and the local extinction of understory birds in the eastern Usambara Mountains, Tanzania. *Conservation Biology*, **5**, 67-78.

Onderdonk, D. & Chapman, C. (2000) Coping with forest fragmentation: the primates of Kibale National Park, Uganda. *International Journal of Primatology*, **21**, 587-611.

Pahl, L.I., Winter, J.W. & Heinsohn, G. (1988) Variation in responses of arboreal marsupials to fragmentation of tropical rainforest in north eastern Australia. *Biological Conservation*, **46**, 71-82.

Pearce, J., Venier, L., Eccles, G., Pedlar, J. & McKenney, D. (2005) Habitat islands, forest edge and spring-active invertebrate assemblages. *Biodiversity and Conservation*, **14**, 2949-2969.

Ramanamanjato, J.-B. (2000) Fragmentation effects on reptile and amphibian diversity in the littoral forest of southeastern Madagascar. *Isolated Vertebrate Communities in the*

Tropics (ed. by G. Rheinwald), pp. 297–308. Zoologisches Forschungsinstitut und Museum A. Koenig, Bonn.

Recher, H.F., Shields, J., Kavanagh, R. & Webb, G. (1987) Retaining remnant mature forest for nature conservation at Eden, New South Wales: a review of theory and practice. *Nature Conservation: The Role of Remnants of Native Vegetation* (ed. by D.A. Saunders, G.W. Arnold, A.A. Burbridge and A.J.M. Hopkins), pp. 177-194. Surrey Beatty & Sons, Chipping Norton.

Rosenblatt, D.L., Heske, E.J., Nelson, S.L., Barber, D.M., Miller, M.A. & MacAllister, B. (1999) Forest fragments in east-central Illinois: islands or habitat patches for mammals? *The American Midland Naturalist*, **141**, 115-123.

Shreeve, T.G. & Mason, C.F. (1980) The Number of Butterfly Species in Woodlands. *Oecologia*, **45**, 414-418.

Soulé, M.E., Wilcox, B.A. & Holtby, C. (1979) Benign neglect: A model of faunal collapse in the game reserves of East Africa. *Biological Conservation*, **15**, 259-272.

Soulé, M.E., Bolger, D.T., Allison, C.A., Wright, J., Sorice, M. & Hill, S. (1988) Reconstructed dynamics of rapid extinctions of chaparral-requiring birds in urban habitat islands. *Conservation Biology*, **2**, 75-92.

Summerville, K.S., Veech, J.A. & Crist, T.O. (2002) Does variation in patch use among butterfly species contribute to nestedness at fine spatial scales? *Oikos*, **97**, 195-204.

Tellería, J.L. & Santos, T. (1995) Effects of forest fragmentation on a guild of wintering passerines: The role of habitat selection. *Biological Conservation*, **71**, 61-67.

Viveiros de Castro, E.B. & Fernandez, F.A.S. (2004) Determinants of differential extinction vulnerabilities of small mammals in Atlantic forest fragments in Brazil. *Biological Conservation*, **119**, 73-80.

Watson, D.M. (2003) Long-term consequences of habitat fragmentation–highland birds in Oaxaca, Mexico. *Biological Conservation*, **111**, 283-303.

Watson, D.M., Mac Nally, R. & Bennett, A.F. (2000) The avifauna of severely fragmented, Buloke Allocasuarina luehmanni woodland in western Victoria, Australia. *Pacific Conservation Biology*, **6**, 46-60.

Weaver, M. & Kellman, M. (1981) The effects of forest fragmentation on woodlot tree biotas in Southern Ontario. *Journal of Biogeography*, **8**, 199-210.

Western, D. & Ssemakula, J. (1981) The future of the savannah ecosystems: ecological islands or faunal enclaves? *African Journal of Ecology*, **19**, 7-19.

Whitcomb, R.F., Robbins, C.S., Lynch, J.F., Whitcomb, B.L., Klimkiewicz, M.K. & Bystrak, D. (1981) Effects of forest fragmentation on avifauna of the Eastern Deciduous Forest. *Forest Island Dynamics in Man-Dominated Landscapes* (ed. by R.L. Burgess and D.M. Sharpe), pp. 125-205. Springer-Verlag, New York.

Woolhouse, M.E.J. (1983) The theory and practice of the species–area effect, applied to the breeding birds of British Woods. *Biological Conservation*, **27**, 315-332.

A2.2 Supplementary results

Table A2.1 Comparison of the number of selected best island species–area relationship models before (BC) and after (AC) applying the Cook's Distance criterion for each of the different versions of the dataset (untransformed, semi-log and log-log). Note that it is the "after" set of models which are the focus of the paper. n=76. Note also that six of the best models changed between the BC and AC analyses for log-log versions of the datasets, but the overall number of best models in each category remains unchanged.

	Untransformed		Semi-Log		Log-Log	
	BC	AC	BC	AC	BC	AC
Simple, linear (positive)	14	14	24	26	24	24
Simple, linear (negative)	1	1	0	0	0	0
Piecewise – continuous (steep-shallow)	17	13	5	5	12	12
Piecewise – continuous (shallow-steep)	0	0	6	5	0	0
Piecewise – zero slope	0	1	11	11	5	5
Piecewise - discontinuous	4	4	6	5	7	7
Non-Significant	40	43	24	24	28	28

Table A2.2 The results of the sensitivity analysis: the best fit ISAR model for 76 habitat island datasets, tallying the best fit models separately for each data transformation. The results presented relate to the model selection after accounting for the initial optimization of the breakpoints and increase in model complexity by increasing the number of model parameters to k=4 for the piecewise continuous model with zero slope, k=5 for the piecewise continuous model, and k=6 for the piecewise discontinuous model, when calculating AICc. The best model for each version of the dataset (i.e. untransformed, semi-log and log-log) was chosen based on AIC_c ($\Delta AIC_c>2$) calculating using the aforementioned number of parameters. These results were tallied following the application of Cook's distance criterion and are thus comparable with Table 1.

	Number of datasets (Percentages in parentheses)		
Model Type	Untransformed	Semi-Log	Log-Log
Linear (positive relationship)	16 (21%)	36 (47%)	32 (42%)
Linear (negative relationship)	1 (1%)	0	0
Piecewise continuous (steep-shallow)	9 (12%)	2 (3%)	9 (12%)
Piecewise continuous (shallow-steep)	0	1 (1%)	0
Piecewise Zero Slope (i.e. flat-steep)	0	5 (7%)	1 (1%)
Piecewise Discontinuous	3 (4%)	4 (5%)	3 (4%)
Non-Significant	47 (62%)	28 (37%)	31 (41%)

Table A2.3 The best fit ISAR model for habitat island datasets according to the three taxonomic groups, tallying the best fit models separately for each data transformation. Studies were grouped into three taxonomic groups: a) vertebrates, b) plants, and c) invertebrates. The best model within a transformation was chosen based on AIC_c (ΔAIC_c>2). These results were tallied following the application of Cook's distance criterion (see 'Materials and Methods').

a) Vertebrates (n=39)

	Number of datasets (Percentages in parentheses)		
Model Type	Untransformed	Semi-Log	Log-Log
Linear (positive relationship)	9 (23%)	15 (38%)	10 (26%)
Linear (negative relationship)	0 (0%)	0 (0%)	0 (0%)
Piecewise continuous (steep-shallow)	10 (26%)	4 (10%)	9 (23%)
Piecewise continuous (shallow-steep)	0 (0%)	2 (5%)	0 (0%)
Piecewise Zero Slope (i.e. flat-steep)	1 (3%)	8 (21%)	4 (10%)
Piecewise Discontinuous	0 (0%)	0 (0%)	3 (8%)
Non-Significant	19 (49%)	10 (26%)	13 (33%)

b) Plants (n=23)

	Number of datasets (Percentages in parentheses)		
Model Type	Untransformed	Semi-Log	Log-Log
Linear (positive relationship)	4 (17%)	7 (30%)	10 (43%)
Linear (negative relationship)	0 (0%)	0 (0%)	0 (0%)
Piecewise continuous (steep-shallow)	4 (17%)	0 (0%)	2 (8%)
Piecewise continuous (shallow-steep)	0 (0%)	2 (8%)	0 (0%)
Piecewise Zero Slope (i.e. flat-steep)	0 (0%)	2 (8%)	0 (0%)
Piecewise Discontinuous	3 (13%)	5 (22%)	4 (17%)
Non-Significant	12 (52%)	7 (30%)	7 (30%)

c) Invertebrates (n=14)

	Number of datasets (Percentages in parentheses)		
Model Type	Untransformed	Semi-Log	Log-Log
Linear (positive relationship)	1 (7%)	4 (29%)	4 (29%)
Linear (negative relationship)	1 (7%)	0 (0%)	0 (0%)
Piecewise continuous (steep-shallow)	3 (21%)	1 (7%)	1 (7%)
Piecewise continuous (shallow-steep)	0 (0%)	1 (7%)	0 (0%)
Piecewise Zero Slope (i.e. flat-steep)	0 (0%)	1 (7%)	1 (7%)
Piecewise Discontinuous	1 (7%)	0 (0%)	0 (0%)
Non-Significant	8 (57%)	7 (50%)	8 (57%)

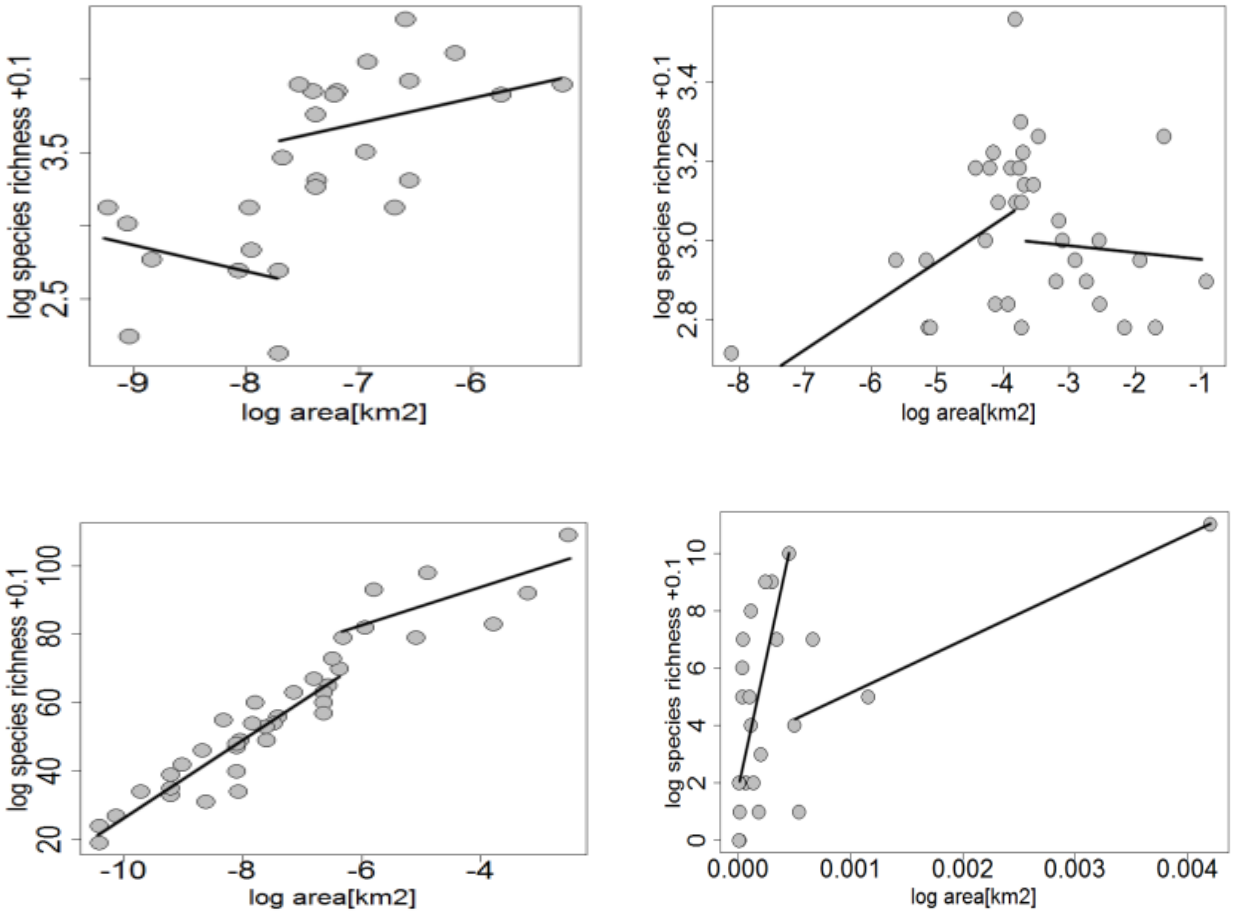


Figure A2.1 A selection of plots depicting discontinuous piecewise models selected as the best for particular datasets. Plot source data (clockwise from top left): Crowe (1979), Levenson (1981; shrub species), Darlington *et al.* (2001), Hattori & Ishida (2000).

Appendix 3

Supporting information: On the form of species–area relationships in habitat islands and true islands (Chapter 5)

Appendix 3.1 Expanded study methods

Empirical dataset sampling methodology

Two fragmented landscapes in Europe were sampled for birds by one of the authors (TJM) during April–July in 2012. The two landscapes were located as follows:

- North France (n=40 fragments; 50° 8'52.11"N, 1°40'37.09"E)
- South Spain (n=43 fragments; 36°11'31.16"N, 5°30'13.61"W)

where the latitude / longitude co-ordinates are the approximate co-ordinates of the centre of each study landscape.

Fragments of varying area (see Table S1 in Appendix S2) were sampled in each landscape using ten-minute point counts of 50m radius. The number of point counts per fragment was proportional to the area of the fragment, and the exact number of point counts was determined using a combination of pilot studies and recommendations from the literature: 1 point count for fragments < 5ha, 2 counts in fragments 5-10ha, 3 counts in fragments 10-15ha, 4 counts in fragments 15-20ha, and so on. Point counts were located systematically, such that there was a minimum of a minimum 200m between counts (Bibby *et al.*, 2000), and to ensure that both the patch core and edge were sampled (birds outside the fragment were not counted). Each fragment was sampled three times, with the exact time of sampling randomised in each instance, and species accumulation curves were constructed to ensure sampling was relatively complete. All birds seen or heard within the point counts were recorded. Individuals that were clearly transient (e.g. flying over the fragment) were removed prior to analysis. Counts took place from dawn to 10am. Island area was measured using ArcGIS (version 10.0; ESRI, 2011).

Bibby, C.J., Burgess, N.D. & Hill, D.A. (2000) *Bird census techniques*, 2nd edn. Academic Press, London.

ESRI (2011) *ArcGIS Desktop: Release 10*. Environmental Systems Research Institute.

3.2 Dataset source information.

Table A3.1 Description of the source papers, including: the taxon, number of islands (No. Isl.), the range in area between the smallest and the largest island (Area range), and the range in species richness between the most species-poor and most species-rich island (Sp. range). The best-fitting ISAR model and the observed shape of this model are also provided. The best model was chosen by comparing 20 candidate ISAR models (see Table 5.1) using an information theoretic approach. The model with the lowest AIC_c was considered as the best-fitting model, if the fit was deemed satisfactory. The best model shape refers to the observed shape of the fit according to the sequential algorithm of Triantis *et al.* (2012). NA refers to datasets in which either none of the candidate models provided a satisfactory fit, or for which AIC_c could not be computed due to too few islands (see 'Materials and Methods'). Certain datasets were obtained from the authors of the source papers, whilst others were supplemented with additional data from the source paper authors. Full references follow the table. * indicates inland water-body island datasets which were added to those collected by Triantis *et al.* (2012).

Dataset	Taxon	No. Isl.	Area range (Ha)	Sp. range	Best model	Model shape
Alvarenga & Pôrto (2007)	Plants	8	493	28	expo	Convex
Alvarenga & Pôrto (2007)	Plants	8	493	19	linear	Linear
Amaral et al. (2012)	Plants	6	398	309	NA	NA
Anciães & Marini (2000)	Vertebrates	5	335	8	NA	NA
Báldi & Kisbenedek (1999)	Invertebrates	27	40	17	epm1	Convex
Baldwin & Bradfield (2007)	Plants	10	62	9	monod	Convex
Baldwin & Bradfield (2007)	Plants	10	23	14	monod	Convex
Baz et al. (1995)	Invertebrates	13	2211	17	monod	Convex
Behle (1978)	Vertebrates	14	92463	45	expo	Convex
Bell & Donnelly (2006)	Vertebrates	9	6	17	linear	Linear
Benassi et al. (2007)	Vertebrates	16	30	60	power	Convex
Benedick et al. (2006)	Invertebrates	8	122907	20	monod	Convex
Bhagwat et al. (2005)	Vertebrates	24	48	34	expo	convex do
Blake & Karr	Vertebrates	12	598	20	koba	Convex

1984						
Bogich et al. (2012)	Invertebrates	12	180	49	negexpo	Convex
Bolger et al. (1991)	Vertebrates	30	102	4	power	Convex
Bolger et al. (1997)	Vertebrates	25	84	7	power	Convex
Boyle & Smith (2010)	Vertebrates	9	99	5	negexpo	Convex
Brashares et al. (2001)	Vertebrates	6	478200	17	NA	NA
Brotons & Herrando (2001)	Vertebrates	51	38	17	koba	Convex
Brown (1971)	Vertebrates	17	301994	11	negexpo	Convex
Brown (1978)	Vertebrates	13	178969	5	expo	Convex
Brown (1978)	Vertebrates	19	301993	10	monod	Convex
Browne (1981)	Vertebrates	12	2571640	103	power	Convex
Browne (1981)	Plants	25	2571880	75	ratio	Convex
Browne (1981)	Invertebrates	13	2571640	30	power	Convex
Burke & Goulet (1998)	Invertebrates	8	2316	44	expo	Convex
Bzdon (2008)	Plants	20	2	92	asymp	convex do
Cabrera-Guzmán & Reynoso (2012)	Vertebrates	6	16	15	NA	NA
Calvillo et al. (2010)	Invertebrates	9	19	24	power	Convex
Carter-Lovejoy (1982)	Vertebrates	14	4	4	linear	Linear
Carter-Lovejoy (1982)	Plants	14	4	24	chapman	Convex
Castelletta et al. (2005)	Vertebrates	6	928	49	NA	NA
Cayuela et al. (2006)	Plants	16	5593	67	linear	Linear
Charles & Ang (2010)	Vertebrates	9	149	12	power	Convex
Cieślak & Dombrowski (1993)	Vertebrates	50	15	37	koba	Convex
Croci et al. (2008)	Vertebrates	13	2	12	monod	Convex
Croci et al. (2008)	Invertebrates	12	2	34	linear	Linear
Croci et al. (2008)	Plants	13	2	28	monod	Convex
Croci et al. (2008)	Vertebrates	10	2	3	linear	Linear

Crooks (2002)	Vertebrates	29	100	5	NA	NA
Crooks et al. (2001)	Vertebrates	30	100	7	koba	Convex
Crowe (1979)	Plants	26	1	60	monod	Convex
Daily & Ehrlich (1995)	Invertebrates	8	224	18	expo	Convex
Dami et al. (2013)	Vertebrates	31	40	79	power	Convex
Darlington et al. (2001)	Invertebrates	25	0	11	expo	Convex
Davis & Jones (1986)	Invertebrates	28	1	13	koba	Convex
Davis et al. (1988)	Vertebrates	28	5845800	25	power	Convex
Davis et al. (1988)	Vertebrates	12	2180500	10	koba	Convex
Dickman (1987)	Vertebrates	9	2	11	weibull3	Sigmoid
Dickman (1987)	Vertebrates	12	7	15	koba	Convex
Dickman (1987)	Vertebrates	8	7	7	linear	Linear
Dinesen et al. (2001)	Vertebrates	18	52100	6	negexpo	Convex
dos Anjos & Boçon (1999)	Vertebrates	12	840	93	gompertz	Sigmoid
dos Anjos (2001)	Vertebrates	5	59	30	NA	NA
dos Santos et al. (2007)	Plants	13	233	63	NA	NA
Edenius & Sjöberg (1997)	Vertebrates	18	176	16	power	Convex
Essl & Dirnböck (2012)	Invertebrates	60	7	12	P1	Convex
Feeley & Terborgh (2006)*	Verts	29	21.2	19	NA	NA
Feeley (2003)*	Verts	26	179.8	41	NA	NA
Fernández-Juricic (2000)	Vertebrates	25	117	18	expo	Convex
Ferraz et al. (2003)	Vertebrates	11	100	42	power	Convex
Filgueiras et al. (2011)	Invertebrates	19	3490	20	linear	Linear
Flaspohler et al. 2010	Vertebrates	18	56	7	koba	Convex
Florencio & Diehl (2006)	Invertebrates	7	9	4	linear	Linear

Ford (1987)	Vertebrates	20	18	22	monod	Convex
Frank & Battisti (2005)	Vertebrates	16	526	23	expo	Convex
Fuentes-Montemayor et al. (2012)	Plants	34	30	13	monod	Convex
Galanes & Thomlinson (2009)	Plants	12	31	21	linear	Linear
Galle (2008)	Invertebrates	15	0	19	koba	Convex
Galli et al. (1976)	Vertebrates	10	24	32	koba	Convex
Ganzhorn et al. (1999)	Vertebrates	7	597	7	linear	Linear
Ganzhorn et al. (1999)	Plants	7	597	12	linear	Linear
Ganzhorn et al. (2000)	Vertebrates	10	647	4	monod	Convex
Gaublomme et al. (2008)	Invertebrates	10	83	35	expo	Convex
Gavish et al. (2012)	Invertebrates	12	3	46	power	Convex
Gavish et al. (2012)	Invertebrates	12	4	44	expo	Convex
Gillespie & Walter (2001)	Vertebrates	7	16384	25	koba	Convex
Gonçalves & Louzada (2005)	Invertebrates	4	13	27	NA	NA
Granados et al. (2001)	Plants	10	1396	432	expo	Convex
Haila et al. (1993)	Vertebrates	13	3	12	linear	Linear
Hall et al. (2002)	Plants	20	78	23	expo	Convex
Hamel et al. (1993)	Vertebrates	32	287	12	linear	Linear
Hanser & Huntly (2006)	Vertebrates	11	3	3	linear	Linear
Hanser & Huntly (2006)	Vertebrates	11	2	4	power	convex do
Hatfield & LeBuhn (2007)	Invertebrates	20	118	7	epm2	Sigmoid
Hatfield & LeBuhn (2007)	Plants	20	118	107	gompertz	Sigmoid
Hatt (1948)*	Plants	17	15102.8	9	NA	NA
Hatt (1948)*	Verts	15	15102.8	137	NA	NA
Hatt (1948)*	Verts	17	15102.8	13	NA	NA
Hatt (1948)*	Verts	12	14963	29	NA	NA
Hatt (1948)*	Verts	17	15102.8	7	NA	NA

Hattori & Ishida (2000)	Plants	38	8	90	weibull3	Convex
Hendrix et al. (2010)	Invertebrates	5	5	15	NA	NA
Hendrix et al. (2010)	Plants	5	5	17	NA	NA
Holbech (2005)	Vertebrates	15	56430	45	expo	Convex
Horváth et al. (2009)	Invertebrates	8	352	38	expo	convex do
Hu et al. (2012)*	Plants	Lake islands	152	130.8	118	NA
Ishida et al. (1998)	Plants	29	16	66	power	Convex
Johnson (1975)	Vertebrates	20	207718	44	power	Convex
Kelt (2000)	Vertebrates	14	123	3	negexpo	Convex
Kitchener et al. (1980)	Vertebrates	23	5085	31	monod	Convex
Kitchener et al. (1980)	Vertebrates	23	5085	11	negexpo	Convex
Kitchener et al. (1980)	Plants	23	5085	251	linear	Linear
Kitchener et al. (1982)	Vertebrates	22	5085	43	expo	Convex
Kiviniemi & Eriksson (2002)	Plants	16	14	36	P1	Convex
Knoll & Penatti (2012)	Invertebrates	5	2256	3	NA	NA
Koh & Sodhi (2004)	Invertebrates	18	1146	26	asymp	Convex
Kratter (1992)	Vertebrates	20	106829	41	power	Convex
Krügel & dos Anjos (2000)	Vertebrates	5	51	36	NA	NA
Langrand (1995)	Vertebrates	8	1249	37	expo	Convex
Lasky & Keitt (2010)	Vertebrates	5	1035	14	NA	NA
Lawesson et al. (1998)	Plants	62	445	157	power	Convex
Leal et al. (2012)	Invertebrates	18	338	38	negexpo	Convex
Lehvävirta et al. (2006)	Invertebrates	15	339	21	P1	NO SHAPE
Lomolino & Davis (1997)	Vertebrates	24	2115900	15	power	Convex
Lomolino et al. (1989)	Vertebrates	27	1112744	15	monod	Convex
Lorrenzetti & Battisti (2007)	Vertebrates	20	302	13	expo	Convex
Lorrenzetti	Plants	20	302	7	expo	Convex

&Battisti (2007)						
Loyn (1987)	Vertebrates	56	1771	59	weibull3	Convex
Lumaret et al. (1997)	Plants	8	0	71	linear	Linear
Lyman (1997) *	Verts	12	108.7	9	NA	NA
Magura et al. (2001)	Invertebrates	15	3954	19	NA	NA
Mahan & O'Connell (2005)	Vertebrates	7	9	4	negexpo	Convex
Maldonado-Coelho & Marini (2003)	Vertebrates	9	380	18	power	Convex
Mamo & Bolen (1999)	Vertebrates	9	81	13	power	Convex
Marini (2001)	Vertebrates	6	222	57	NA	NA
Marshall et al. (2010)	Vertebrates	22	52626	4	koba	Convex
Martínez-Morales (2005)	Vertebrates	13	16288	51	linear	Linear
Matthiae & Stearns (1981)	Vertebrates	22	40	9	epm2	Sigmoid
McCollin (1993)	Vertebrates	16	14	20	monod	Convex
Metzger et al. (1997)	Plants	15	73	97	monod	Convex
Meynard & Quinn (2008)	Vertebrates	9	60650	8	negexpo	Convex
Miyashita et al. (1998)	Invertebrates	7	27	18	monod	Convex
Miyashita et al. (1998)	Invertebrates	9	15	20	power	Convex
Mohd-Azlan & Lawes (2011)	Vertebrates	13	594	23	monod	Convex
Mohd-Azlan & Lawes (2011)	Plants	13	594	24	NA	NA
Natuhara & Imai (1999)	Vertebrates	28	43	13	linear	Linear
Neame et al. (2013)	Invertebrates	19	32	22	expo	Convex
Newmark (1986)	Vertebrates	24	2065300	29	expo	Convex
Newmark (1991)	Vertebrates	10	521	24	koba	Convex
Nores (1995)	Vertebrates	9	358700	46	linear	Linear
Nufio et al. (2011)	Invertebrates	13	36	19	power	Convex
Nyeko (2009)	Invertebrates	6	140	17	NA	NA

Oliver et al. (2011)	Vertebrates	20	5908	125	expo	Convex
Page et al. (2010)	Plants	11	45	25	negexpo	Convex
Page et al. (2010)	Plants	11	45	10	linear	Linear
Page et al. (2010)	Plants	11	45	13	monod	Convex
Page et al. (2010)	Plants	11	45	34	linear	Linear
Pahl et al. (1988)	Vertebrates	11	72	4	power	Convex
Pearce et al. (2005)	Invertebrates	10	11	21	negexpo	Convex
Peltzer et al. (2003)	Vertebrates	8	1	17	power	Convex
Pineda & Halffter (2004)	Vertebrates	5	56	5	NA	NA
Raheem et al. (2008)	Invertebrates	21	29227	23	power	Convex
Ramanamanjato (2000)	Vertebrates	11	447	19	power	Convex
Ramanamanjato (2000)	Vertebrates	11	447	26	power	Convex
Rau & Gantz (2001)	Vertebrates	18	341	12	expo	Convex
Recher et al. (1987)	Plants	22	11	4	negexpo	Convex
Recher et al. (1987)	Vertebrates	22	11	5	power	Convex
Ribas et al. (2005)	Invertebrates	18	296	64	gompertz	Sigmoid
Ribeiro et al. (2012)	Invertebrates	10	72	22	linear	Linear
Ríos et al. (2011)	Vertebrates	5	78	12	NA	NA
Rosenblatt et al. (1999)	Vertebrates	10	598	6	gompertz	Convex
Ruiz-Gutiérrez et al. (2008)	Vertebrates	7	260	82	power	Convex
Sánchez-Gallen (2010)	Plants	7	637	21	power	convex do
Scanlan (1981)	Plants	40	31	40	power	Convex
Scanlan (1981)	Plants	26	774	66	monod	Convex
Schieck et al. (1995)	Vertebrates	5	2480	4	NA	NA
Schieck et al. (1995)	Vertebrates	9	496	8	power	convex do
Schieck et al.	Vertebrates	7	356	5	linear	Linear

(1995)						
Schoereder et al. (2004)	Invertebrates	17	90	64	epm1	Sigmoid
Shreeve & Mason (1980)	Invertebrates	22	173	21	expo	Convex
Silva & Pôrto (2009)	Plants	10	2606	44	linear	Linear
Silva (2001)	Vertebrates	14	6	5	linear	Linear
Silva et al. (2003)	Vertebrates	11	40	7	power	convex do
Simberloff & Martin (1991)	Vertebrates	35	101	31	negexpo	Convex
Smith et al. (1996)	Vertebrates	26	174	16	linear	Linear
Soulé et al. (1988)	Vertebrates	37	103	7	koba	Convex
Sroka & Finch (2006)	Invertebrates	10	325	11	expo	convex do
Stiles & Scheiner (2010)	Plants	22	8763	27	koba	Convex
Struebig et al. (2008)	Vertebrates	15	11308	18	negexpo	Convex
Suarez et al. (1998)	Invertebrates	40	102	20	koba	Convex
Summerville et al. (2002)	Invertebrates	25	0	17	linear	Linear
Tellería & Santos (1995)	Vertebrates	27	350	8	power	Convex
This study	Vertebrates	40	150	35	ratio	Convex
This study	Vertebrates	42	110	28	ratio	Convex
Tonn & Magnuson (1982)	Vertebrates	18	88	11	negexpo	Convex
Uehara-Prado et al. (2007)	Invertebrates	5	161	16	NA	NA
Usher & Keiller (1998)	Invertebrates	18	31	62	power	Convex
Usher et al. (1993)	Invertebrates	28	10	19	monod	Convex
Usher et al. (1993)	Invertebrates	28	10	12	linear	Linear
Vallan (2000)	Vertebrates	7	1250	17	power	Convex
Vignoli et al. (2009)	Vertebrates	44	1591	19	P2	Sigmoid
Viveiros de Castro & Fernandez (2004)	Vertebrates	8	12	6	negexpo	Convex
Wang et al.	Verts	42	1288.9	58	NA	NA

(2010)*						
Wang et al. (2010)*	Verts	42	1288.9	5	NA	NA
Watling & Donnelly (2008)	Vertebrates	24	7	11	power	Convex
Watling & Donnelly (2008)	Vertebrates	24	7	6	NA	NA
Watson & Peterson (1999)	Vertebrates	31	129900	126	power	Convex
Watson (2003)	Vertebrates	17	159244	74	asymp	Convex
Watson et al. (2000)	Vertebrates	27	465	15	linear	Linear
Watson et al. (2009)	Vertebrates	30	464	51	power	Convex
Watts & Larivière (2004)	Invertebrates	4	76	34	NA	NA
Weaver & Kellman (1981)	Plants	10	7	10	koba	Convex
Wethered & Lawes (2005)	Vertebrates	9	215	26	linear	Linear
Wethered & Lawes (2005)	Vertebrates	9	273	11	power	Convex
Whitcomb et al. (1981)	Vertebrates	27	356	11	expo	Convex
Wijesinghe (2012)	Vertebrates	15	10855	5	linear	Linear
Wilcox et al. (1986)	Invertebrates	18	292900	51	power	Convex
Williams & Pearson (1997)	Vertebrates	22	169429	59	P2	Sigmoid
Williams & Pearson (1997)	Vertebrates	22	169429	21	koba	Convex
Williams & Pearson (1997)	Vertebrates	22	169429	39	NA	NA
Williams & Pearson (1997)	Vertebrates	22	169429	24	linear	Linear
Willson et al. (1994)	Vertebrates	10	350	12	linear	Linear
Wilson et al. (2007)	Vertebrates	12	683	3	power	Convex
Winter & Faaborg (1999)	Vertebrates	13	1053	5	power	Convex
Woolhouse (1983)	Vertebrates	30	36	20	power	Convex
Zimmerman & Bierregaard	Vertebrates	7	499	32	koba	Convex

(1986)

Source paper references

Alvarenga, L.D.P. & Pôrto, K.C. (2007) Patch size and isolation effects on epiphytic and epiphyllous bryophytes in the fragmented Brazilian Atlantic forest. *Biological Conservation*, **134**, 415-427.

Amaral, D.D., Vieira, I.C.G., Salomão, R.P., Almeida, S.S. & Jardim, M.A.G. (2012) The status of conservation of urban forests in eastern Amazonia. *Brazilian Journal of Biology*, **72**, 257-265.

Anciães, M. & Marini, M.Â. (2000) The effects of fragmentation on fluctuating asymmetry in passerine birds of Brazilian tropical forests. *Journal of Applied Ecology*, **37**, 1013-1028.

Báldi, A. & Kisbenedek, T. (1999) Orthopterans in small steppe patches: an investigation for the best-fit model of the species-area curve and evidences for their non-random distribution in the patches. *Acta Oecologica*, **20**, 125-132.

Baldwin, L.K. & Bradfield, G.E. (2007) Bryophyte responses to fragmentation in temperate coastal rainforests: a functional group approach. *Biological Conservation*, **136**, 408-422.

Baz, A. & Garcia-Boyero, A. (1995) The effects of forest fragmentation on butterfly communities in central Spain. *Journal of Biogeography*, **22**, 129-140.

Behle, W.H. (1978) Avian biogeography of the Great Basin and intermountain region. *Great Basin Naturalist Memoirs*, **2**, 55-80.

Bell, K.E. & Donnelly, M.A. (2006) Influence of forest fragmentation on community structure of frogs and lizards in northeastern Costa Rica. *Conservation Biology*, **20**, 1750-1760.

Benassi, G., Battisti, C. & Luiselli, L. (2007) Area effect on bird species richness of an archipelago of wetland fragments in Central Italy. *Community Ecology*, **8**, 229-237.

Benedick, S., Hill, J.K., Mustaffa, N., Chey, V.K., Maryati, M., Searle, J.B., Schilthuizen, M. & Hamer, K.C. (2006) Impacts of rain forest fragmentation on butterflies in northern Borneo: species richness, turnover and the value of small fragments. *Journal of Applied Ecology*, **43**, 967-977.

Bhagwat, S.A., Kushalappa, C.G., Williams, P.H. & Brown, N.D. (2005) The role of informal protected areas in maintaining biodiversity in the Western Ghats of India. *Ecology and Society*, **10**, 8-48.

Blake, J.G. & Karr, J.R. (1984) Species composition of bird communities and the conservation benefit of large versus small forests. *Biological Conservation*, **30**, 173-187.

Bogich, T.L., Barker, G.M., Mahlfeld, K., Climo, F., Green, R. & Balmford, A. (2012) Fragmentation, grazing and the species–area relationship. *Ecography*, **35**, 224-231.

Bolger, D.T., Allison, C.A. & Soulé, M.E. (1991) Occurrence patterns of bird species in habitat fragments: sampling, extinction, and nested species subsets. *The American Naturalist*, **137**, 155-166.

Bolger, D.T., Allison, C.A., Sauvajot, R.M., Potenza, P., McCalvin, C., Tran, D., Mazzoni, S. & Soulé, M.E. (1997) Response of rodents to habitat fragmentation in coastal southern California. *Ecological Applications*, **7**, 552-563.

Boyle, S.A. & Smith, A.T. (2010) Can landscape and species characteristics predict primate presence in forest fragments in the Brazilian Amazon? *Biological Conservation*, **143**, 1134-1143.

Brashares, J.S., Arcese, P. & Sam, M.K. (2001) Human demography and reserve size predict wildlife extinction in West Africa. *Proceedings of the Royal Society B: Biological Sciences*, **268**, 2473-2478.

Brotons, L. & Herrando, S. (2001) Reduced bird occurrence in pine forest fragments associated with road proximity in a Mediterranean agricultural area. *Landscape and Urban Planning*, **57**, 77-89.

Brown, J.H. (1971) Mammals on mountaintops: nonequilibrium insular biogeography. *The American Naturalist*, **105**, 467-478.

Brown, J.H. (1978) The theory of island biogeography and the distribution of boreal birds and mammals. *Great Basin Naturalist Memoirs*, **2,** 209-227.

Browne, R.A. (1981) Lakes as islands: biogeographic distribution, turnover rates, and species composition in the lakes of central New York. *Journal of Biogeography*, **8**, 75-83.

Burke, D. & Goulet, H. (1998) Landscape and area effects on beetle assemblages in Ontario. *Ecography*, **21**, 472-479.

Bzdon, G. (2008) Gravel pits as habitat islands: floristic diversity and vegetation analysis. *Polish Journal of Ecology*, **56**, 239-250.

Cabrera-Guzmán, E. & Reynoso, V. (2012) Amphibian and reptile communities of rainforest fragments: minimum patch size to support high richness and abundance. *Biodiversity and Conservation*, **21**, 3243-3265.

Calvillo, L.M., Ramírez, V.M., Parra-Tabla, V. & Navarro, J. (2010) Bee diversity in a fragmented landscape of the Mexican neotropic. *Journal of Insect Conservation*, **14**, 323-334.

Carter-Lovejoy, S.H. (1982) The relation between species numbers and island characteristics for habitat islands in a volcanic landscape. *Great Basin Naturalist*, **42**, 113-119.

Castelletta, M., Thiollay, J.-M. & Sodhi, N.S. (2005) The effects of extreme forest fragmentation on the bird community of Singapore Island. *Biological Conservation*, **121**, 135-155.

Cayuela, L., Golicher, D.J., Benayas, J.M.R., González-Espinosa, M. & RamÍrez-Marcial, N. (2006) Fragmentation, disturbance and tree diversity conservation in tropical montane forests. *Journal of Applied Ecology*, **43**, 1172-1181.

Charles, J. & Ang, B. (2010) Non-volant small mammal community responses to fragmentation of kerangas forests in Brunei Darussalam. *Biodiversity and Conservation*, **19**, 543-561.

Cieślak, M. & Dombrowski, A. (1993) The effect of forest size on breeding bird communities. *Acta Ornithologica*, **27**, 97-111.

Croci, S., Butet, A., Georges, A., Aguejdad, R. & Clergeau, P. (2008) Small urban woodlands as biodiversity conservation hot-spot: a multi-taxon approach. *Landscape Ecology*, **23,** 1171-1186.

Crooks, K.R. (2002) Relative sensitivities of mammalian carnivores to habitat fragmentation. *Conservation Biology*, **16**, 488-502.

Crooks, K.R., Suarez, A.V., Bolger, D.T. & Soulé, M.E. (2001) Extinction and colonization of birds on habitat islands. *Conservation Biology*, **15**, 159-172.

Crowe, T.M. (1979) Lots of weeds: insular phytogeography of vacant urban lots. *Journal of Biogeography*, **6,** 169-181.

Daily, G.C. & Ehrlich, P.R. (1995) Preservation of biodiversity in small rainforest patches: rapid evaluations using butterfly trapping. *Biodiversity and Conservation*, **4**, 35-55.

Dami, F.D., Mwansat, G.S. & Manu, S.A. (2013) The effects of forest fragmentation on species richness on the Obudu Plateau, south-eastern Nigeria. *African Journal of Ecology*, **51**, 32-36.

Darlington, J.P.E.C., Kaib, M. & Brand, R. (2001) Termites (Isoptera) in forest remnants and forest islands in the Shimba Hills National Reserve, Coastal Province of Kenya. *Sociobiology*, **37**, 527-538.

Davis, B.N.K. & Jones, P.E. (1986) Insects on isolated colonies of common rock-rose *Helianthemum chamaecistus*. *Ecological Entomology*, **11**, 267-281.

Davis, R., Dunford, C. & Lomolino, M.V. (1988) Montane mammals of the American Southwest: the possible influence of post-Pleistocene colonization. *Journal of Biogeography*, **15**, 841-848.

Dickman, C.R. (1987) Habitat fragmentation and vertebrate species richness in an urban environment. *Journal of Applied Ecology*, **24**, 337-351.
Dinesen, L., Lehmberg, T., Rahner, M.C. & Fjeldså, J. (2001) Conservation priorities for the forests of the Udzungwa Mountains, Tanzania, based on primates, duikers and birds. *Biological Conservation*, **99**, 223-236.
dos Anjos, L. (2001) Bird communities in five Atlantic forest fragments in southern Brazil. *Ornitologia Neotropical*, **12**, 11-27.
dos Anjos, L. & Boçon, R. (1999) Bird communities in natural forest patches in southern Brazil. *The Wilson Bulletin*, **111**, 397-414.
dos Santos, K., Kinoshita, L.S. & dos Santos, F.A.M. (2007) Tree species composition and similarity in semideciduous forest fragments of southeastern Brazil. *Biological Conservation*, **135**, 268-277.
Edenius, L. & Sjöberg, K. (1997) Distribution of birds in natural landscape mosaics of old-growth forests in northern Sweden: relations to habitat area and landscape context. *Ecography*, **20**, 425-431.
Essl, F. & Dirnböck, T. (2012) What determines Orthoptera species distribution and richness in temperate semi-natural dry grassland remnants? *Biodiversity and Conservation*, **21**, 2525-2537.
Feeley, K. (2003) Analysis of avian communities in Lake Guri, Venezuela, using multiple assembly rule models. *Oecologia*, **137**, 104-113.
Feeley, K.J. & Terborgh, J.W. (2006) Habitat fragmentation and effects of herbivore (howler monkey) abundances on bird species richness. *Ecology*, **87**, 144-150.
Fernández-Juricic, E. (2000) Bird community composition patterns in urban parks of Madrid: the role of age, size and isolation. *Ecological Research*, **15**, 373-383.
Ferraz, G., Russell, G.J., Stouffer, P.C., Bierregaard, R.O., Pimm, S.L. & Lovejoy, T.E. (2003) Rates of species loss from Amazonian forest fragments. *Proceedings of the National Academy of Sciences*, **100**, 14069-14073.
Filgueiras, B.K.C., Iannuzzi, L. & Leal, I.R. (2011) Habitat fragmentation alters the structure of dung beetle communities in the Atlantic Forest. *Biological Conservation*, **144**, 362-369.
Flaspohler, D.J., Giardina, C.P., Asner, G.P., Hart, P., Price, J., Lyons, C.K.a. & Castaneda, X. (2010) Long-term effects of fragmentation and fragment properties on bird species richness in Hawaiian forests. *Biological Conservation*, **143**, 280-288.
Florencio, D.F. & Diehl, E. (2006) Termitofauna (Insecta, Isoptera) em remanescentes de floresta estacional semidecidual em São Leopoldo, Rio Grande do Sul, Brasil. *Revista Brasileira de Entomologia*, **50**, 505-511.
Ford, H.A. (1987) Bird communities on habitat islands in England. *Bird Study*, **34**, 205 - 218.
Frank, B. & Battisti, C. (2005) Area effect on bird communities, guilds and species in a highly fragmented forest landscape of central Italy. *Italian Journal of Zoology*, **72**, 297-304.
Fuentes-Montemayor, E., Goulson, D., Cavin, L., Wallace, J.M. & Park, K.J. (2012) Factors influencing moth assemblages in woodland fragments on farmland: Implications for woodland management and creation schemes. *Biological Conservation*, **153**, 265-275.
Galanes, I. & Thomlinson, J. (2009) Relationships between spatial configuration of tropical forest patches and woody plant diversity in northeastern Puerto Rico. *Plant Ecology*, **201**, 101-113.
Galle, R. (2008) The effect of a naturally fragmented landscape on the spider assemblages. *North-Western Journal of Zoology*, **4**, 61-71.
Galli, A.E., Leck, C.F. & Forman, R.T.T. (1976) Avian distribution patterns in forest islands of different sizes in central New Jersey. *The Auk*, **93**, 356-364.
Ganzhorn, J.U., Fietz, J., Rakotovao, E., Schwab, D. & Zinner, D. (1999) Lemurs and the regeneration of dry deciduous forest in Madagascar. *Conservation Biology*, **13**, 794-804.

Ganzhorn, J.U., Goodman, S.M., Ramanamanjato, J.-B., Ralison, J., Rakotondravony, D. & Rakotosamimanana, B. (2000) Effects of fragmentation and assessing minimum viable populations of lemurs in Madagascar. *Isolated vertebrate communities in the tropics* (ed. by G. Rheinwald), pp. 265-272. Zoologisches Forschungsinstitut und Museum A. Koenig, Bonn.

Gaublomme, E., Hendrickx, F., Dhuyvetter, H. & Desender, K. (2008) The effects of forest patch size and matrix type on changes in carabid beetle assemblages in an urbanized landscape. *Biological Conservation*, **141**, 2585-2596.

Gavish, Y., Ziv, Y. & Rosenzweig, M.L. (2012) Decoupling fragmentation from habitat loss for spiders in patchy agricultural landscapes. *Conservation Biology*, **26**, 150-159.

Gillespie, T.W. & Walter, H. (2001) Distribution of bird species richness at a regional scale in tropical dry forest of Central America. *Journal of Biogeography*, **28**, 651-662.

Gonçalves, T.T. & Louzada, J.N.C. (2005) Estratificação vertical de coleópteros carpófilos (Insecta:Coleoptera) em fragmentos florestais do sul do Estado de Minas Gerais, Brasil. *Ecología Austral*, **15**, 101-110.

Granados, M., O'Kennon, R.J. & Benz, B.F. (2001) Plant species–area relationships in ten central Texas protected natural areas. *SIDA*, **19**, 1061-1072.

Haila, Y., Hanski, I.K. & Raivio, S. (1993) Turnover of breeding birds in small forest fragments: the "sampling" colonization hypothesis corroborated. *Ecology*, **74**, 714-725.

Hall, J.M., Gillespie, T.W., Richardson, D. & Reader, S. (2002) Fragmentation of Florida scrub in an urban landscape. *Urban Ecosystems*, **6**, 243-243.

Hamel, P.B., Smith, W.P. & Wahl, J.W. (1993) Wintering bird populations of fragmented forest habitat in the central basin, Tennessee. *Biological Conservation*, **66**, 107-115.

Hanser, S.E. & Huntly, N.J. (2006) The biogeography of small mammals of fragmented sagebrush-steppe landscapes. *Journal of Mammalogy*, **87**, 1165-1174.

Hatfield, R.G. & LeBuhn, G. (2007) Patch and landscape factors shape community assemblage of bumble bees, Bombus spp. (Hymenoptera: Apidae), in montane meadows. *Biological Conservation*, **139**, 150-158.

Hatt, R. (1948) *Island life: a study of land vertebrates of the islands of eastern Lake Michigan.* Cranbrook Institute of Science, Michigan.

Hattori, T. & Ishida, H. (2000) Relationship between species diversity, species composition and forest area of fragmented lucidophyllous forests in central Miyazaki Prefecture. *Japanese Journal of Ecology*, **50**, 221-234.

Hendrix, S., Kwaiser, K. & Heard, S. (2010) Bee communities (Hymenoptera: Apoidea) of small Iowa hill prairies are as diverse and rich as those of large prairie preserves. *Biodiversity and Conservation*, **19**, 1699-1709.

Hoffmann, M.D. & Dodson, S.I. (2005) Land use, primary productivity, and lake area as descriptors of zooplankton diversity. *Ecology*, **86**, 255-261.

Holbech, L.H. (2005) The implications of selective logging and forest fragmentation for the conservation of avian diversity in evergreen forests of south-west Ghana. *Bird Conservation International*, **15**, 27-52.

Horváth, R., Magura, T., Szinetár, C. & Tóthmérész, B. (2009) Spiders are not less diverse in small and isolated grasslands, but less diverse in overgrazed grasslands: a field study (East Hungary, Nyírség). *Agriculture, Ecosystems & Environment*, **130**, 16-22.

Hu, G., Wu, J., Feeley, K.J., Xu, G. & Yu, M. (2012) The effects of landscape variables on the species-area relationship during late-stage habitat fragmentation. *PLOS ONE*, **7**, e43894.

Ishida, H., Hattori, T., Takeda, Y. & Kodate, S. (1998) Relationship between species richness or species composition and area of fragmented lucidophyllous forests in southeastern Hyogo Prefecture. *Japanese Journal of Ecology*, **48**, 1-16.

Johnson, N.K. (1975) Controls of number of bird species on montane islands in the Great Basin. *Evolution*, 29, 545-567.

Kelt, D.A. (2000) Small mammal communities in rainforest fragments in central southern Chile. *Biological Conservation*, **92**, 345-358.
Kitchener, D.J., Chapman, A., Muir, B.G. & Palmer, M. (1980a) The conservation value for mammals of reserves in the western Australian wheatbelt. *Biological Conservation*, **18**, 179-207.
Kitchener, D.J., Dell, J., Muir, B.G. & Palmer, M. (1982) Birds in Western Australian wheatbelt reserves—implications for conservation. *Biological Conservation*, **22**, 127-163.
Kitchener, D.J., Chapman, A., Dell, J., Muir, B.G. & Palmer, M. (1980b) Lizard assemblage and reserve size and structure in the Western Australian wheatbelt — some implications for conservation. *Biological Conservation*, **17**, 25-62.
Kiviniemi, K. & Eriksson, O. (2002) Size-related deterioration of semi-natural grassland fragments in Sweden. *Diversity and Distributions*, **8**, 21-29.
Knoll, F. & Penatti, N. (2012) Habitat fragmentation effects on the orchid bee communities in remnant forests of southeastern Brazil. *Neotropical Entomology*, **41**, 355-365.
Koh, L.P. & Sodhi, N.S. (2004) Importance of reserves, fragments, and parks for butterfly conservation in a tropical urban landscape. *Ecological Applications*, **14**, 1695-1708.
Kratter, A.W. (1992) Montane avian biogeography in southern California and Baja California. *Journal of Biogeography*, **19**, 269-283.
Krügel, M.M. & dos Anjos, L. (2000) Bird communities in forest remnants in the city of Maringá, Paraná State, southern Brazil. *Ornitologia Neotropical*, **11**, 315-330.
Langrand, O. (1995) *The effects of forest fragmentation on bird species in Madagascar: a case study from Ambohitantely Forest Reserve on the central high plateau.* University of Natal, Pietermaritzburg.
Lasky, J.R. & Keitt, T.H. (2010) Abundance of Panamanian dry-forest birds along gradients of forest cover at multiple scales. *Journal of Tropical Ecology*, **26**, 67-78.
Lawesson, J.E., de Blust, G., Grashof, C., Firbank, L., Honnay, O., Hermy, M., Hobitz, P. & Jensen, L.M. (1998) Species diversity and area-relationships in Danish beech forests. *Forest Ecology and Management*, **106**, 235-245.
Leal, I., Filgueiras, B., Gomes, J., Iannuzzi, L. & Andersen, A. (2012) Effects of habitat fragmentation on ant richness and functional composition in Brazilian Atlantic forest. *Biodiversity and Conservation*, **21**, 1687-1701.
Lehvävirta, S., Kotze, D., Niemelä, J., Mäntysaari, M. & O'Hara, B. (2006) Effects of fragmentation and trampling on carabid beetle assemblages in urban woodlands in Helsinki, Finland. *Urban Ecosystems*, **9**, 13-26.
Lomolino, M.V. & Davis, R. (1997) Biogeographic scale and biodiversity of mountain forest mammals of western North America. *Global Ecology and Biogeography Letters*, **6**, 57-76.
Lomolino, M.V., Brown, J.H. & Davis, R. (1989) Island biogeography of montane forest mammals in the American Southwest. *Ecology*, **70**, 180-194.
Lorenzetti, E. & Battisti, C. (2007) Nature reserve selection on forest fragments in a suburban landscape (Rome, Central Italy): indications from a set of avian species. *Landscape Research*, **32**, 57-78.
Loyn, R.H. (1987) Effects of patch area and habitat on bird abundances, species numbers and tree health in fragmented Victorian forests. *Nature conservation: the role of remnants of native vegetation* (ed. by D.A. Saunders, G.W. Arnold, A.A. Burbridge and A.J.M. Hopkins), pp. 65-77. Surrey Beatty and Sons, Chipping Norton.
Lumaret, R., Guillerm, J.-L., Maillet, J. & Verlaque, R. (1997) Plant species diversity and polyploidy in islands of natural vegetation isolated in extensive cultivated lands. *Biodiversity and Conservation*, **6**, 591-613.
Lyman, A.J. (1997) Rapid decline of small mammal diversity in monsoon evergreen forest fragments in Thailand. *Tropical forest remnants: ecology, management and conservation of fragmented*

communities (ed. by W.F. Laurance and R.O. Bierregaard), pp. 220-240. University of Chicago Press, Chicago, US.
Magura, T., Ködöböcz, V. & Tóthmérész, B. (2001) Effects of habitat fragmentation on carabids in forest patches. *Journal of Biogeography*, **28**, 129-138.
Mahan, C.G. & O'Connell, T.J. (2005) Small mammal use of suburban and urban parks in central Pennsylvania. *Northeastern Naturalist*, **12**, 307-314.
Maldonado-Coelho, M. & Marini, M.Â. (2000) Effects of forest fragment size and successional stage on mixed-species bird flocks in southeastern Brazil. *The Condor*, **102**, 585-594.
Mamo, L.B. & Bolen, E.G. (1999) Effects of area, isolation, and landscape on the avifauna of Carolina Bays. *Journal of Field Ornithology*, **70**, 310-320.
Marini, M.Â. (2001) Effects of forest fragmentation on birds of the Cerrado region, Brazil. *Bird Conservation International*, **11**, 13-25.
Marshall, A.R., Jørgensbye, H.I.O., Rovero, F., Platts, P.J., White, P.C.L. & Lovett, J.C. (2010) The species–area relationship and confounding variables in a threatened monkey community. *American Journal of Primatology*, **72**, 325-336.
Martínez-Morales, M.A. (2005) Landscape patterns influencing bird assemblages in a fragmented neotropical cloud forest. *Biological Conservation*, **121**, 117-126.
Matthiae, P.E. & Stearns, F. (1981) Mammals in forests islands in southeastern Wisconsin. *Forest island dynamics in man-dominated landscapes.* (ed. by R.L. Burgess and D.M. Sharpe), pp. 55-66. Springer-Verlag, New York.
McCollin, D. (1993) Avian distribution patterns in a fragmented wooded landscape (North Humberside, U.K.): the role of between-patch and within-patch structure. *Global Ecology and Biogeography Letters*, **3**, 48-62.
Metzger, J.P., Bernacci, L.C. & Goldenberg, R. (1997) Pattern of tree species diversity in riparian forest fragments of different widths (SE Brazil). *Plant Ecology*, **133**, 135-152.
Meynard, C.N. & Quinn, J.F. (2008) Bird metacommunities in temperate South American forest: vegetation structure, area, and climate effects. *Ecology*, **89**, 981-990.
Miyashita, T., Shinkai, A. & Chida, T. (1998) The effects of forest fragmentation on web spider communities in urban areas. *Biological Conservation*, **86**, 357-364.
Mohd-Azlan, J. & Lawes, M.J. (2011) The effect of the surrounding landscape matrix on mangrove bird community assembly in north Australia. *Biological Conservation*, **144**, 2134-2141.
Natuhara, Y. & Imai, C. (1999) Prediction of species richness of breeding birds by landscape-level factors of urban woods in Osaka Prefecture, Japan. *Biodiversity and Conservation*, **8**, 239-253.
Neame, L.A., Griswold, T. & Elle, E. (2013) Pollinator nesting guilds respond differently to urban habitat fragmentation in an oak-savannah ecosystem. *Insect Conservation and Diversity*, **6**, 57-66.
Newmark, W.D. (1986) Species-area relationship and its determinants for mammals in western North American national parks. *Biological Journal of the Linnean Society*, **28**, 83-98.
Newmark, W.D. (1991) Tropical forest fragmentation and the local extinction of understory birds in the eastern Usambara Mountains, Tanzania. *Conservation Biology*, **5**, 67-78.
Nores, M. (1995) Insular biogeography of birds on mountain-tops in north western Argentina. *Journal of Biogeography*, **22**, 61-70.
Nufio, C., McClenahan, J. & Deane Bowers, M. (2011) Grasshopper response to reductions in habitat area as mediated by subfamily classification and life history traits. *Journal of Insect Conservation*, **15**, 409-419.
Nyeko, P. (2009) Dung beetle assemblages and seasonality in primary forest and forest fragments on agricultural landscapes in Budongo, Uganda. *Biotropica*, **41**, 476-484.

Oliver, A.J., Hong-Wa, C., Devonshire, J., Olea, K.R., Rivas, G.F. & Gahl, M.K. (2011) Avifauna richness enhanced in large, isolated urban parks. *Landscape and Urban Planning*, **102**, 215-225.

Page, N., Qureshi, Q., Rawat, G. & Kushalappa, C. (2010) Plant diversity in sacred forest fragments of Western Ghats: a comparative study of four life forms. *Plant Ecology*, **206**, 237-250.

Pahl, L.I., Winter, J.W. & Heinsohn, G. (1988) Variation in responses of arboreal marsupials to fragmentation of tropical rainforest in north eastern Australia. *Biological Conservation*, **46**, 71-82.

Pearce, J., Venier, L., Eccles, G., Pedlar, J. & McKenney, D. (2005) Habitat islands, forest edge and spring-active invertebrate assemblages. *Biodiversity and Conservation*, **14**, 2949-2969.

Peltzer, P.M., Lajmanovich, R.C. & Beltzer, A.H. (2003) The effects of habitat fragmentation on amphibian species richness in the floodplain of the Middle Parana River, Argentina. *Herpetological Journal*, **13**, 95-98.

Pineda, E. & Halffter, G. (2004) Species diversity and habitat fragmentation: frogs in a tropical montane landscape in Mexico. *Biological Conservation*, **117**, 499-508.

Raheem, D.C., Naggs, F., Preece, R.C., Mapatuna, Y., Kariyawasam, L. & Eggleton, P. (2008) Structure and conservation of Sri Lankan land-snail assemblages in fragmented lowland rainforest and village home gardens. *Journal of Applied Ecology*, **45**, 1019-1028.

Ramanamanjato, J.-B. (2000) Fragmentation effects on reptile and amphibian diversity in the littoral forest of southeastern Madagascar. *Isolated vertebrate communities in the tropics.* (ed. by G. Rheinwald), pp. 297-308. Zoologisches Forschungsinstitut und Museum A. Koenig, Bonn.

Rau, J. & Gantz, A. (2001) Fragmentation of the native rainforest of southern Chile: effects of the area and shape on the biodiversity of birds. *Boletin de la Sociedad de Biología de Concepción*, **72**, 109 -119

Recher, H.F., Shields, J., Kavanagh, R. & Webb, G. (1987) Retaining remnant mature forest for nature conservation at Eden, New South Wales: a review of theory and practice. *Nature conservation: the role of remnants of native vegetation* (ed. by D.A. Saunders, G.W. Arnold, A.A. Burbridge and A.J.M. Hopkins), pp. 177-194. Surrey Beatty & Sons, Chipping Norton.

Ribas, C.R., Sobrinho, T.G., Schoereder, J.H., Sperber, C.F., Lopes-Andrade, C. & Soares, S.M. (2005) How large is large enough for insects? Forest fragmentation effects at three spatial scales. *Acta Oecologica*, **27**, 31-41.

Ribeiro, D., Batista, R., Prado, P., Brown, K. & Freitas, A. (2012) The importance of small scales to the fruit-feeding butterfly assemblages in a fragmented landscape. *Biodiversity and Conservation*, **21**, 811-827.

Ríos, S.S., Lloyd, H. & Valdés-Velásquez, A. (2011) Bird species richness, diversity and abundance in Polylepis woodlands, Huascaran biosphere reserve, Peru. *Studies on Neotropical Fauna and Environment*, **46**, 69-76.

Rosenblatt, D.L., Heske, E.J., Nelson, S.L., Barber, D.M., Miller, M.A. & MacAllister, B. (1999) Forest Fragments in east-central Illinois: islands or habitat patches for mammals? *The American Midland Naturalist*, **141**, 115-123.

Ruiz-Gutiérrez, V., Gavin, T.A. & Dhondt, A.A. (2008) Habitat fragmentation lowers survival of a tropical forest bird. *Ecological Applications*, **18**, 838-846.

Sánchez-Gallen, I., Álvarez-Sánchez, F.J. & Benítez-Malvido, J. (2010) Structure of the advanced regeneration community in tropical rain forest fragments of Los Tuxtlas, Mexico. *Biological Conservation*, **143**, 2111-2118.

Scanlan, M.J. (1981) Biogeography of forest plants in the prairie-forest ecotone in western Minnesota. *Forest island dynamics in man-dominated landscapes* (ed. by R.L. Burgess and D.M. Sharpe), pp. 97-124. Springer-Verlag, New York

Schieck, J.I.M., Lertzman, K.E.N., Nyberg, B. & Page, R. (1995) Effects of patch size on birds in old-growth montane forests. *Conservation Biology*, **9**, 1072-1084.

Schoereder, J.H., Sobrinho, T.G., Ribas, C.R. & Campos, R.B.F. (2004) Colonization and extinction of ant communities in a fragmented landscape. *Austral Ecology*, 29, 391-398.

Shreeve, T.G. & Mason, C.F. (1980) The number of butterfly species in woodlands. *Oecologia*, **45**, 414-418.

Silva, M. (2001) Abundance, diversity, and community structure of small mammals in forest fragments in Prince Edward Island National Park, Canada. *Canadian Journal of Zoology*, **79**, 2063-2071.

Silva, M. & Pôrto, K. (2009) Effect of fragmentation on the community structure of epixylic bryophytes in Atlantic Forest remnants in the northeast of Brazil. *Biodiversity and Conservation*, **18**, 317-337.

Silva, M., Hartling, L.A., Field, S.A. & Teather, K. (2003) The effects of habitat fragmentation on amphibian species richness of Prince Edward Island. *Canadian Journal of Zoology*, **81**, 563-573.

Simberloff, D. & Martin, J.L. (1991) Nestedness of insular avifaunas: simple summary statistics masking complex species patterns. *Ornis Fennica*, **68**, 178-192.

Smith, G.T., Arnold, G.W., Sarre, S., Abensperg-Traun, M. & Steven, D.E. (1996) The effect of habitat fragmentation and livestock grazing on animal communities in remnants of gimlet *Eucalyptus salubris* woodland in the western Australian wheatbelt. II. Lizards. *Journal of Applied Ecology*, **33**, 1302-1310.

Soulé, M.E., Bolger, D.T., Allison, C.A., Wright, J., Sorice, M. & Hill, S. (1988) Reconstructed dynamics of rapid extinctions of chaparral-requiring birds in urban habitat islands. *Conservation Biology*, **2**, 75-92.

Sroka, K. & Finch, O.-D. (2006) Ground beetle diversity in ancient woodland remnants in north-western Germany (Coleoptera, Carabidae). *Journal of Insect Conservation*, **10**, 335-350.

Stiles, A. & Scheiner, S.M. (2010) A multi-scale analysis of fragmentation effects on remnant plant species richness in Phoenix, Arizona. *Journal of Biogeography*, **37**, 1721-1729.

Struebig, M.J., Kingston, T., Zubaid, A., Mohd-Adnan, A. & Rossiter, S.J. (2008) Conservation value of forest fragments to Palaeotropical bats. *Biological Conservation*, **141**, 2112-2126.

Suarez, A.V., Bolger, D.T. & Case, T.J. (1998) Effects of fragmentation and invasion on native ant communities in coastal southern California. *Ecology*, **79**, 2041-2056.

Summerville, K.S., Veech, J.A. & Crist, T.O. (2002) Does variation in patch use among butterfly species contribute to nestedness at fine spatial scales? *Oikos*, **97**, 195-204.

Swengel, A. & Swengel, S. (2010) The butterfly fauna of Wisconsin bogs: lessons for conservation. *Biodiversity and Conservation*, **19**, 3565-3581.

Tellería, J.L. & Santos, T. (1995) Effects of forest fragmentation on a guild of wintering passerines: The role of habitat selection. *Biological Conservation*, **71**, 61-67.

Tonn, W.M. & Magnuson, J.J. (1982) Patterns in the species composition and richness of fish assemblages in northern Wisconsin lakes. *Ecology*, **63**, 1149-1166.

Uehara-Prado, M., Brown, K.S. & Freitas, A.V.L. (2007) Species richness, composition and abundance of fruit-feeding butterflies in the Brazilian Atlantic Forest: comparison between a fragmented and a continuous landscape. *Global Ecology and Biogeography*, **16**, 43-54.

Usher, M.B. & Keiller, S.W.J. (1998) The macrolepidoptera of farm woodlands: determinants of diversity and community structure. *Biodiversity and Conservation*, **7**, 725-748.

Usher, M.B., Field, J.P. & Bedford, S.E. (1993) Biogeography and diversity of ground-dwelling arthropods in farm woodlands. *Biodiversity Letters*, **1**, 54-62.

Vallan, D. (2000) Influence of forest fragmentation on amphibian diversity in the nature reserve of Ambohitantely, highland Madagascar. *Biological Conservation*, **96**, 31-43.

Vignoli, L., Mocaer, I., Luiselli, L. & Bologna, M.A. (2009) Can a large metropolis sustain complex herpetofauna communities? An analysis of the suitability of green space fragments in Rome. *Animal Conservation*, **12**, 456-466.

Viveiros de Castro, E.B. & Fernandez, F.A.S. (2004) Determinants of differential extinction vulnerabilities of small mammals in Atlantic forest fragments in Brazil. *Biological Conservation*, **119**, 73-80.

Wang, Y., Bao, Y., Yu, M., Xu, G. & Ding, P. (2010) Nestedness for different reasons: the distributions of birds, lizards and small mammals on islands of an inundated lake. *Diversity and Distributions*, **16**, 862-873.

Watling, J.I. & Donnelly, M.A. (2008) Species richness and composition of amphibians and reptiles in a fragmented forest landscape in northeastern Bolivia. *Basic and Applied Ecology*, **9**, 523-532.

Watson, D.M. (2003) Long-term consequences of habitat fragmentation–highland birds in Oaxaca, Mexico. *Biological Conservation*, **111**, 283-303.

Watson, D.M. & Peterson, A.T. (1999) Determinants of diversity in a naturally fragmented landscape: humid montane forest avifaunas of Mesoamerica. *Ecography*, **22**, 582-589.

Watson, D.M., Mac Nally, R. & Bennett, A.F. (2000) The avifauna of severely fragmented, Buloke 'Allocasuarina Luehmanni' woodland in western Victoria, Australia. *Pacific Conservation Biology*, **6**, 46-60.

Watson, J.E.M., Watson, A.W.T., Fischer, J., Carter Ingram, J. & Whittaker, R.J. (2009) Using nestedness and species-accumulation analyses to strengthen a conservation plan for littoral forest birds in south-eastern Madagascar. *International Journal of Biodiversity and Conservation*, **1**, 67-80.

Watts, C.H. & Larivière, M.-C. (2004) The importance of urban reserves for conserving beetle communities: a case study from New Zealand. *Journal of Insect Conservation*, **8**, 47-58.

Weaver, M. & Kellman, M. (1981) The effects of forest fragmentation on woodlot tree biotas in southern Ontario. *Journal of Biogeography*, **8**, 199-210.

Wethered, R. & Lawes, M.J. (2005) Nestedness of bird assemblages in fragmented Afromontane forest: the effect of plantation forestry in the matrix. *Biological Conservation*, **123**, 125-137.

Whitcomb, R.F., Robbins, C.S., Lynch, J.F., Whitcomb, B.L., Klimkiewicz, M.K. & Bystrak, D. (1981) Effects of forest fragmentation on avifauna of the Eastern Deciduous Forest. *Forest island dynamics in man-dominated landscapes* (ed. by R.L. Burgess and D.M. Sharpe), pp. 125-205. Springer-Verlag, New York.

Wijesinghe, M.R. (2012) Predicting effects of rainforest fragmentation from live trapping studies of small mammals in Sri Lanka. *Journal of Threatened Taxa*, **4**, 2629–2636.

Wilcox, B.A., Murphy, D.D., Ehrlich, P.R. & Austin, G.T. (1986) Insular biogeography of the montane butterfly faunas in the Great Basin: comparison with birds and mammals. *Oecologia*, **69**, 188-194.

Williams, S.E. & Pearson, R.G. (1997) Historical rainforest contractions, localized extinctions and patterns of vertebrate endemism in the rainforests of Australia's wet tropics. *Proceedings of the Royal Society B: Biological Sciences*, **264**, 709-716.

Willson, M.F., De Santo, T.L., Sabag, C. & Armesto, J.J. (1994) Avian communities of fragmented south-temperate rainforests in Chile. *Conservation Biology*, **8**, 508-520.

Wilson, J.W., van Aarde, R.J. & van Rensburg, B.J. (2007) Effects of habitat fragmentation on bird communities of sand forests in southern Mozambique. *Ostrich*, **78**, 37-42.

Winter, M. & Faaborg, J. (1999) Patterns of area sensitivity in grassland-nesting birds. *Conservation Biology*, **13**, 1424-1436.

Woolhouse, M.E.J. (1983) The theory and practice of the species-area effect, applied to the breeding birds of British woods. *Biological Conservation*, **27**, 315-332.

Zimmerman, B.L. & Bierregaard, R.O. (1986) Relevance of the equilibrium-theory of island biogeography and species area relations to conservation with a case from Amazonia. *Journal of Biogeography*, **13**, 133-143.

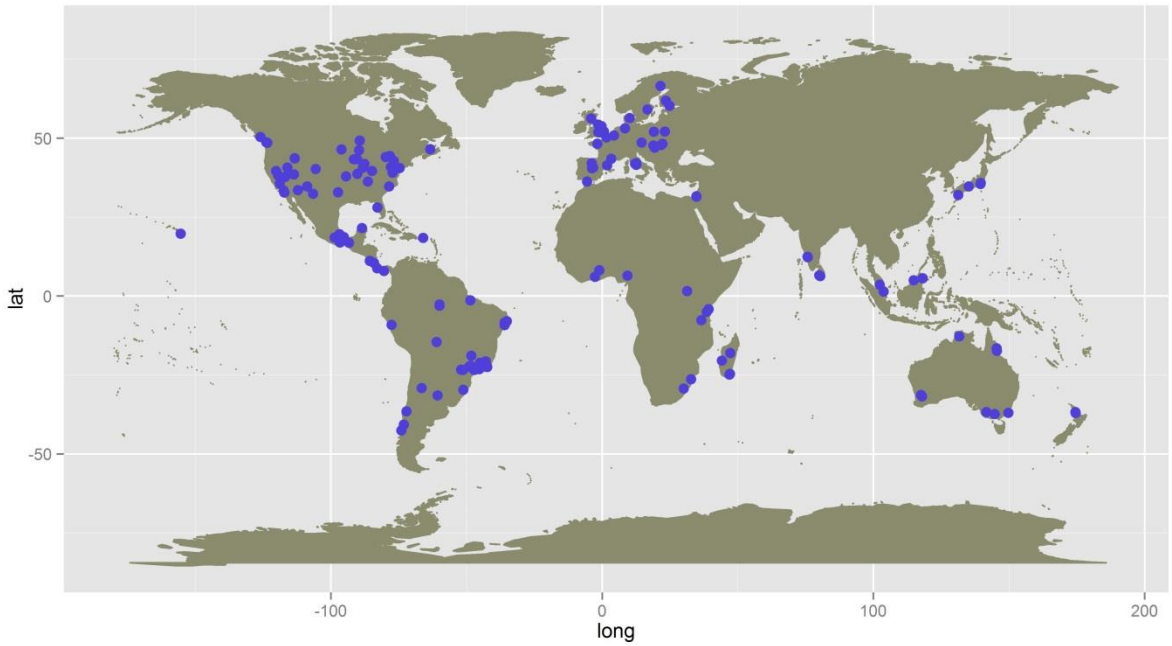


Figure A3.1 A map of the 207 datasets (blue dots) included in this study. Multiple datasets based in the same location are represented by a single blue dot. Where possible we took the coordinates from the centre of the study extent. However, in a small number of papers only a general area was listed (e.g. East Central Illinois), and in these instances we simply used the centre of this area.

Data quality issues

One of the issues with any species–area relationship (SAR) study (or similar synthetic or meta-analytical study) is how to ensure that the sampling methodology chosen and the level of sampling effort are appropriate for the study system under focus. Almost all SAR studies rely on sampling (i.e. total counts of individuals are very rare in the empirical literature) and thus it is a given that there will be some degree of uncertainty surrounding the data gathered. Thus, meta-analyses and synthetic analyses based on published SAR data are also affected by sampling issues.

There are two main potential data quality issues in SAR studies, and by extension SAR meta-analyses. First, an inappropriate sampling method may result in data that generate or contain spurious patterns. Second, variables other than area (e.g. disturbance) are not controlled for. In regards to the first problem, the literature is replete with examples of how sampling effort can influence the form of various macroecological patterns. For example, many species abundance distribution models are based on a particular sampling theory (e.g. Etienne, 2005). The exact influence of low sampling effort on the SAR is less well studied, although it can be surmised that low-intensity sampling is more likely to under-represent the richness of larger islands due to the relatively greater sampling effort needed. As such, the right hand side of the SAR curve may be artificially lowered, favouring particular models over others. One way of addressing this issue may be to only include datasets from studies in which the collection of species lists on islands is the primary methodological aim of the study. By this means, one can be certain that the original authors put a good deal of thought into how the data were to be collected, and that their methods have been peer-reviewed as broadly being fit for the purpose. This is not always the case in ecological meta-analyses. For example, Whittaker (2010) argued that many datasets used in productivity–richness meta-analyses were generated as a by-product of the analyses of the source paper, i.e. the generation of these data was not

the primary aim of the study. In such cases, the data are much more likely to contain problems. Thus, in the present study we attempted to ensure we did not include data that were generated as a by-product of sampling designed for other purposes. Additionally, where possible we tried to determine whether studies used proportional sampling or not. For example, Watson *et al.* (2009; one of our datasets) used a proportional sampling methodology in combination with rarefaction methods to determine when sampling effort was sufficient. Unfortunately, such information is not always clearly provided in source papers. It is possible that others attempting to replicate our study may make different decisions on data suitability but we believe that this is not likely to affect more than a handful of the decisions on dataset screening and that the results presented here should therefore be robust to this particular potential source of bias.

In regards to the second problem, in previous work we have found that confounding variables can influence the performance of certain piecewise regression ISAR models (Matthews *et al.*, 2014). However, this issue is hard to eliminate, as very few habitat island systems contain no confounding variables; by their very nature they are disturbed systems and thus disturbance and isolation etc. will vary between islands within a landscape. It is also the case that many potential confounding variables are not presented in source papers, or were not measured to begin with, prohibiting a systematic analysis of a given set of key variables of potential interest. Partly this is due to the fact that there are a very large number of potential confounding variables in any study system, and workers have limited resources and have framed their research with slightly differing goals. Our approach to this problem was therefore to avoid including studies in which a confounding variable (e.g. isolation) was the primary variable of interest, i.e. over and above area. Furthermore, we have no reason to believe that the inclusion of confounding variables would be biased towards any particular taxon or island type.

Additional references

Etienne, R.S. (2005) A new sampling formula for neutral biodiversity. *Ecology Letters*, **8**, 253–260.

Matthews, T.J., Steinbauer, M.J., Tzirkalli, E., Triantis, K.A. & Whittaker, R.J. (2014) Thresholds and the species–area relationship: a synthetic analysis of habitat island datasets. *Journal of Biogeography*, **41**, 1018–1028.

Watson, J.E.M., Watson, A.W.T., Fischer, J., Ingram, J.C. & Whittaker, R.J. (2009) Using nestedness and species-accumulation analyses to strengthen a conservation plan for littoral forest birds in south-eastern Madagascar. *International Journal of Biodiversity and Conservation*, **1**, 67–80.

Whittaker, R.J. (2010) Meta-analyses and mega-mistakes: calling time on meta-analysis of the species richness–productivity relationship. *Ecology*, **91**, 2522–2533.

Appendix 3.3 Supplementary results

Table A3.2 The results of the robustness analysis for 40 habitat island datasets. The datasets were randomly chosen from our complete set of 207 datasets using a weighting system in which datasets with fewer islands had a higher weighting than datasets with many islands. To test the robustness of our model selection results we used a jacknife procedure. For a given dataset we fitted 18 ISAR models to the complete dataset and defined the best-fitting model (Best model column) as that with the lowest AIC_c value. We then iteratively removed a data point and re-ran the model selection, noting the best-fitting model. The removed data point was re-instated and a different point removed, and the process repeated until the model selection had been run for all possible outcomes/trials. We then calculated the percentage of trials in which the best fitting model matched the best-fitting model for the whole dataset (i.e. all data points; % same as best), and the percentage of trials in which the best-fitting model shape matched the best-fitting model shape for the whole dataset (% same best shape). The Beta-P and Weibull-4 models were not used in this analysis as the fitting process was computationally intensive. The best model names refer to the codes in Table 5.1. Dataset information and source paper references are provided in Appendix A3.2.

Dataset	Best model	No. islands	% same as best	% same best shape
Alvarenga & Pôrto (2007)	expo	8	38	100
Boyle & Smith (2010)	negexpo	9	100	100
Calvillo *et al.* (2010)	power	9	100	100
Charles & Ang (2010)	power	9	78	78
Daily & Ehrlich (1995)	expo	8	0	88
dos Anjos & Boçon (1999)	gompertz	12	83	83
Edenius & Sjöberg (1997)	power	18	83	94
Flaspohler *et al.* 2010	koba	18	61	100
Florencio & Diehl (2006)	linear	7	100	100
Galli *et al.* (1976)	koba	10	30	100
Ganzhorn *et al.* (1999)	linear	7	57	57
Ganzhorn *et al.* (2000)	monod	10	90	100
Gaublomme *et al.* (2008)	expo	10	20	60
Gavish *et al.* (2012)	power	12	42	83
Hanser & Huntly (2006)	linear	11	100	100
Kiviniemi & Eriksson (2002)	P1	16	100	100
Langrand (1995)	expo	8	88	100
Mahan & O'Connell (2005)	negexpo	7	86	100
Maldonado-Coelho & Marini (2003)	power	9	44	44
Mamo & Bolen (1999)	power	9	67	78
McCollin (1993)	monod	16	81	100
Meynard & Quinn (2008)	negexpo	9	56	89
Miyashita *et al.* (1998)	monod	7	100	100
Neame *et al.* (2013)	expo	19	68	89
Nores (1995)	linear	9	89	89
Pearce *et al.* (2005)	negexpo	10	40	100
Peltzer *et al.* (2003)	power	8	63	88
Raheem *et al.* (2008)	power	21	100	100

Ramanamanjato (2000)	power	11	91	91
Ramanamanjato (2000)	power	11	27	73
Rau & Gantz (2001)	expo	18	67	100
Ribeiro *et al.* (2012)	linear	10	100	100
Rosenblatt *et al.* (1999)	gompertz	10	50	90
Ruiz-Gutiérrez *et al.* (2008)	power	7	86	100
Silva *et al.* (2003)	power	11	100	100
Tonn & Magnuson (1982)	negexpo	18	67	78
Vallan (2000)	power	7	100	100
Wethered & Lawes (2005)	linear	9	89	89
Willson *et al.* (1994)	linear	10	70	70
Zimmerman & Bierregaard (1986)	koba	7	57	100

Table A3.3 Parameter estimates for a set of the most parsimonious linear models modelling: (a) the *z* value of the power (log–log model) and (b) the value of the *c* parameter, for all 207 habitat island datasets (i.e. including parameter estimates from non-significant model fits). The predictor variables included: the latitude of the study site (Lat.), area of the smallest island (Amin), area scale (Ascale; the area of the largest island in a dataset divided by the area of the smallest island), the number of islands (Ni.), the minimum and maximum species richness values for a dataset (Smin & Smax), the taxon (Tax.), and the habitat island type (Typ.). The best model (i.e. lowest AIC_c) and all models within ΔAIC_c of < 2 of the best model are given in each instance. The weight of evidence, calculated by summing the Akaike weights of all the models in which a variable was included, is also given. + indicates a significant effect of taxon or habitat type. NA indicates that a variable was not included in a model. (L) indicates predictor variables that were log-transformed. The AIC_c weight and the delta AIC_c (ΔAIC_c) for each model selection are also presented.

c) *z* value

	Lat.	Ascale(L)	Amin(L)	Ni (L)	Smax (L)	Smin (L)	Tax	Typ.	ΔAIC_c	$wAIC_c$
1	NA	-0.01	NA	NA	0.13	-0.12	+	NA	0	0.14
2	<-0.01	-0.01	NA	NA	0.13	-0.13	+	NA	0.65	0.10
3	<-0.01	-0.01	NA	NA	0.12	-0.12	NA	NA	1.92	0.05
	NA	-0.01	NA	NA	0.12	-0.12	NA	NA	1.95	0.05
Weight of evidence	0.45	0.86	0.27	0.28	1	1	0.64	0.25		

d) *c* value

	Lat.	Ascale(L)	Amin(L)	Ni (L)	Smax (L)	Smin (L)	Tax	Typ.	ΔAIC_c	$wAIC_c$
1	NA	-0.07	-0.21	NA	0.32	0.83	NA	+	0	0.37
2	NA	-0.07	-0.2	NA	0.3	0.83	+	+	1.68	0.16
Weight of evidence	0.25	0.98	1	0.25	0.99	1	0.30	0.98		

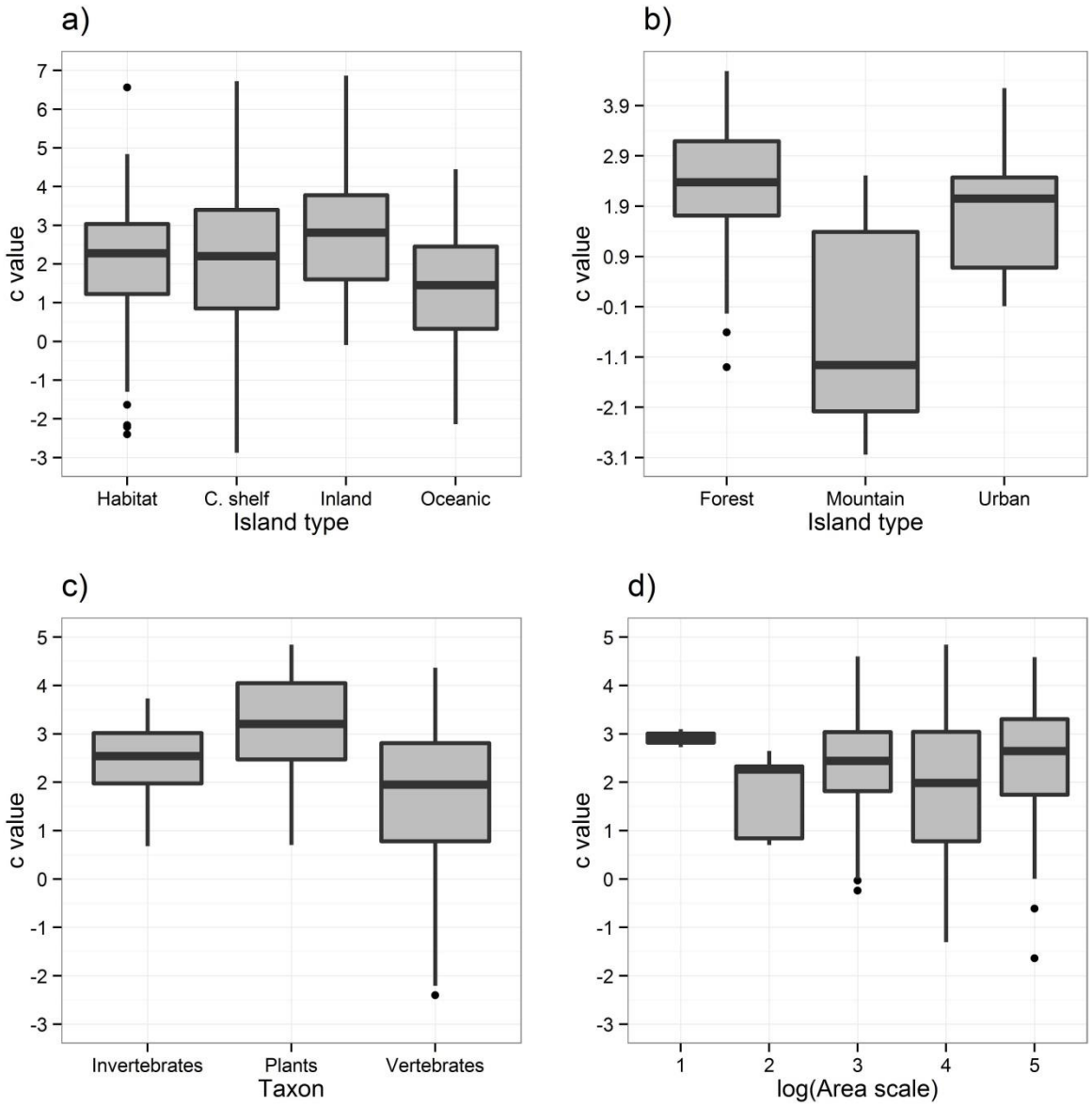


Figure A3.2 Variation in the *c* parameter of the power (log–log) model across: (a) all island types, (b) habitat island types, (c) taxa, and (d) area scale (log transformed). NB Values in (b)–(d) were calculated using only habitat island datasets. Only *c* values from datasets with significant *z* values ($P < 0.05$) were included. In (a) there were 135 habitat island, 125 oceanic island, 58 inland water-body ('Inland'), and 277 continental shelf island datasets ('C. shelf'). In (b) there were 78 forest, 12 mountaintop and 21 urban habitat island datasets. In (c) there were 27 invertebrate, 21 plant and 87 vertebrate datasets. Area scale was calculated as the area of the largest island in a dataset divided by the area of the smallest island. Area scales larger than 5 have been omitted. The box plots display the median (thick black line), the first and third quartiles (thin black box). The whiskers extend from the hinge to the highest value that is within 1.5 multiplied by the inter quartile range, of the hinge.

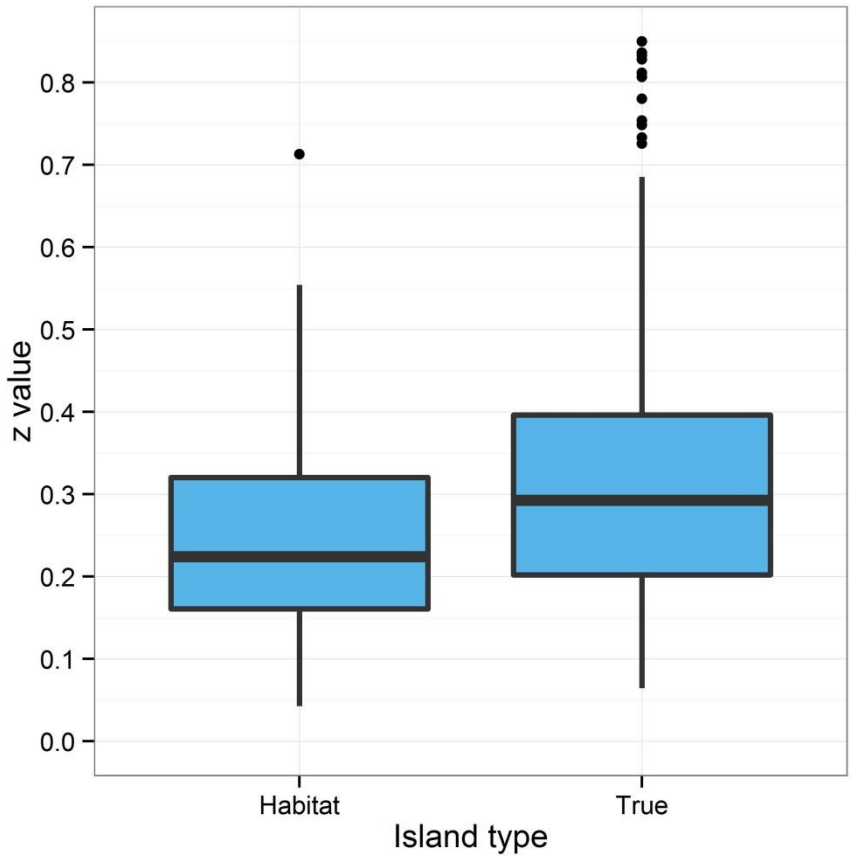


Figure A3.3 Variation in the *z* parameter of the power (log–log) model across habitat islands and true islands. Only *z* values from datasets with significant *z* values ($P < 0.05$) were included. For reasons of clarity, the box plots were constructed after omitting the small number of *z* values < 0. For habitat island datasets this equated to 132 datasets, while there were 460 true island datasets. True islands comprised 125 oceanic island, 58 inland water-body, and 277 continental shelf island datasets. The box plots display the median (thick black line), the first and third quartiles (thin black box). The whiskers extend from the hinge to the highest value that is within 1.5 multiplied by the inter quartile range, of the hinge. Outliers are indicated by solid dots.

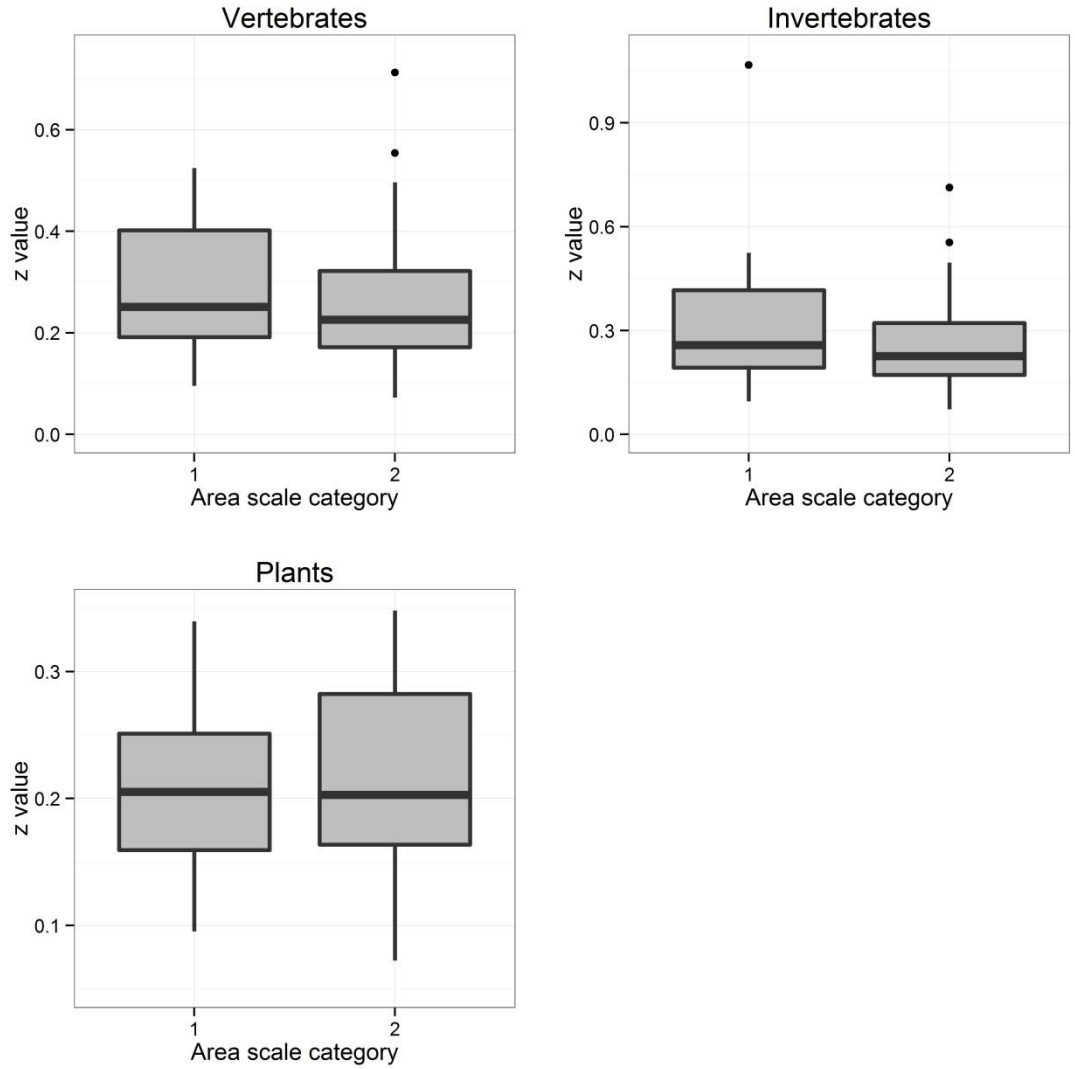


Figure A3.4 Variation in the *z* value of the power (log–log) species–area relationship across different area scales (log-transformed), for three taxa. Area scale was calculated as the area of the largest island in a dataset divided by the area of the smallest island. Only datasets with significant *z* values ($P < 0.05$) were included. For reasons of clarity, the box plots were constructed after omitting the small number of *z* values < 0. There were 86 vertebrate, 26 invertebrate and 20 plant datasets. Due to low sample size, scales 1, 2 and 3 have been combined (area scale category 1), as have all scales larger than 3 (area scale category 2).

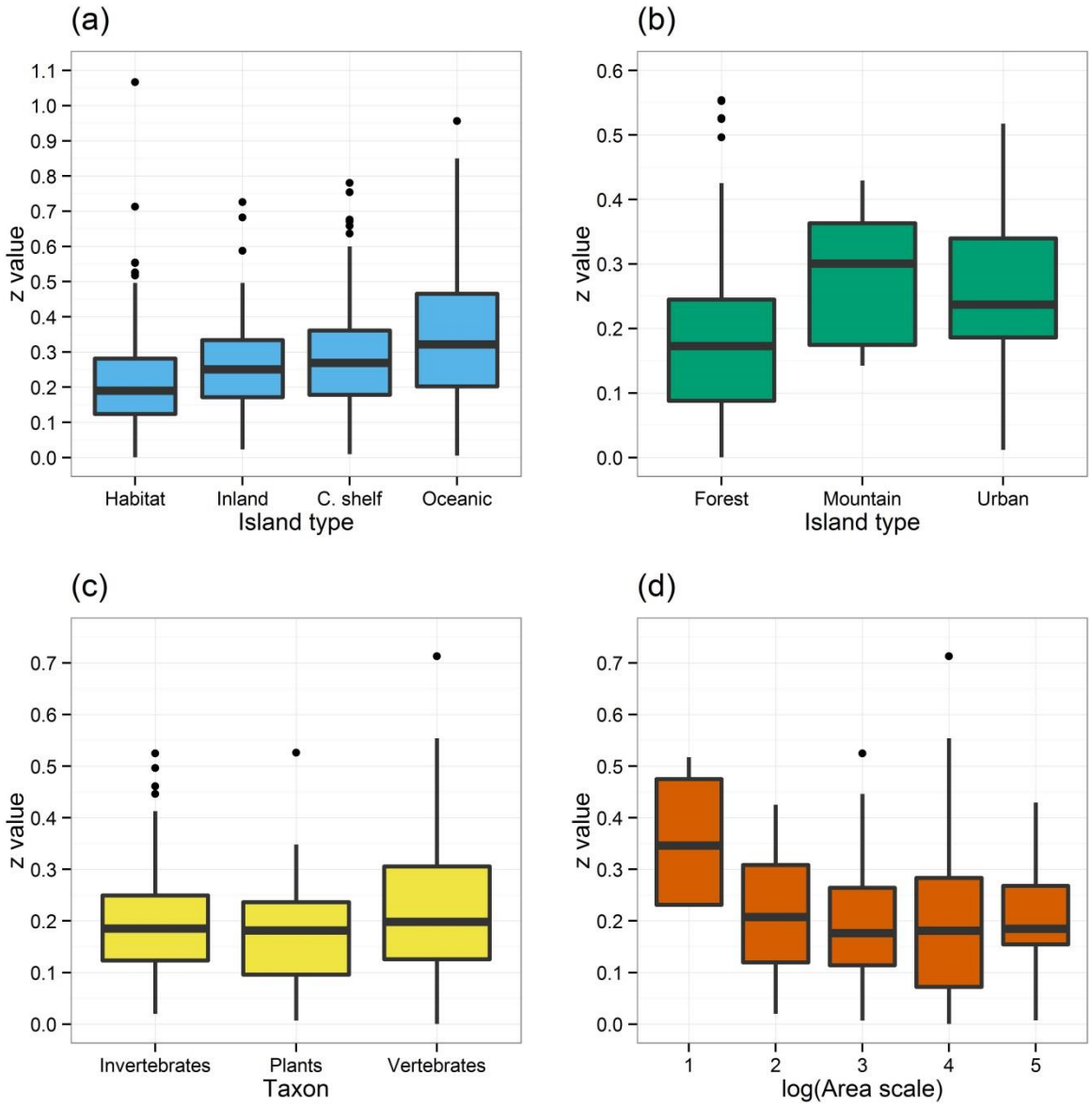


Figure A3.5 Variation in the *z* parameter of the power (log–log) model across (a) all island types, (b) habitat island types, (c) taxa, and (d) area scale (log transformed). (b)-(d) were calculated using only habitat island datasets. Both datasets with significant ($P < 0.05$) and non-significant ($P > 0.05$) *z* values were included. For reasons of clarity, the box plots were constructed after omitting the small number of *z* values < 0 (all subsequent statistics were performed using the full set of data). Thus, in (a) there were 185 habitat island, 188 oceanic island, 66 inland water-body ('Inland'), and 350 continental shelf island datasets ('C. shelf'). Area scale was calculated as the area of the largest island in a dataset divided by the area of the smallest island. Area scales larger than 5 have been omitted. The box plots display the median (thick black line), the first and third quartiles (thin black box). The whiskers extend from the hinge to the highest value that is within 1.5 multiplied by the inter quartile range, of the hinge. Outliers are indicated by solid dots.

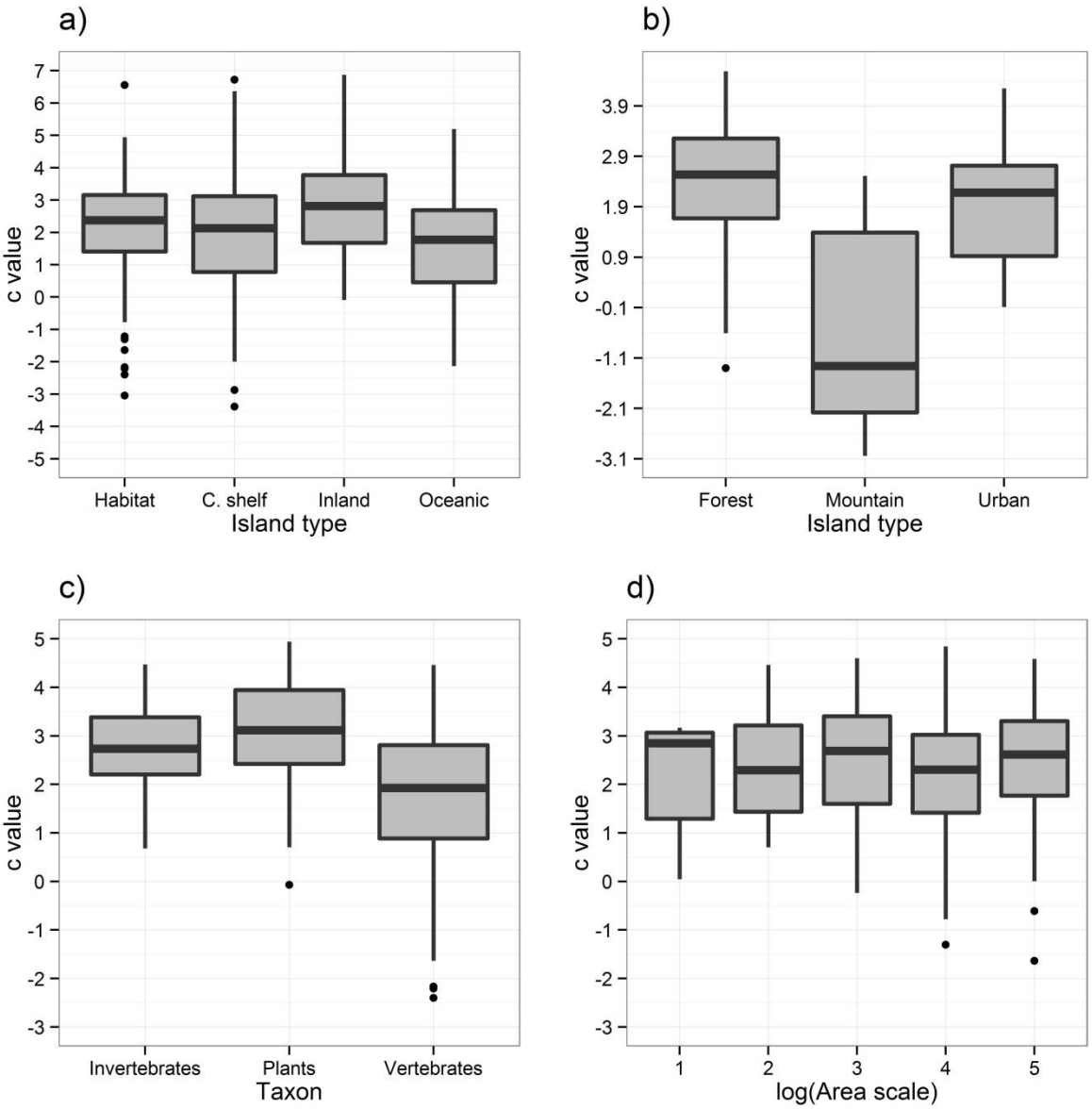


Figure A3.6 Variation in the *c* parameter of the power (log–log) model across (a) all island types, (b) habitat island types, (c) taxa, and (d) area scale (log transformed). (b)-(d) were calculated using only habitat island datasets. Both datasets with significant ($P < 0.05$) and non-significant ($P > 0.05$) *z* values were included. In (a) there were 207 habitat island, oceanic island, 66 inland water-body ('Inland'), and 353 continental shelf island datasets ('C. shelf'). Area scale was calculated as the area of the largest island in a dataset divided by the area of the smallest island. Area scales larger than 5 have been omitted. The box plots display the median (thick black line), the first and third quartiles (thin black box). The whiskers extend from the hinge to the highest value that is within 1.5 multiplied by the inter quartile range, of the hinge. Outliers are indicated by solid dots.

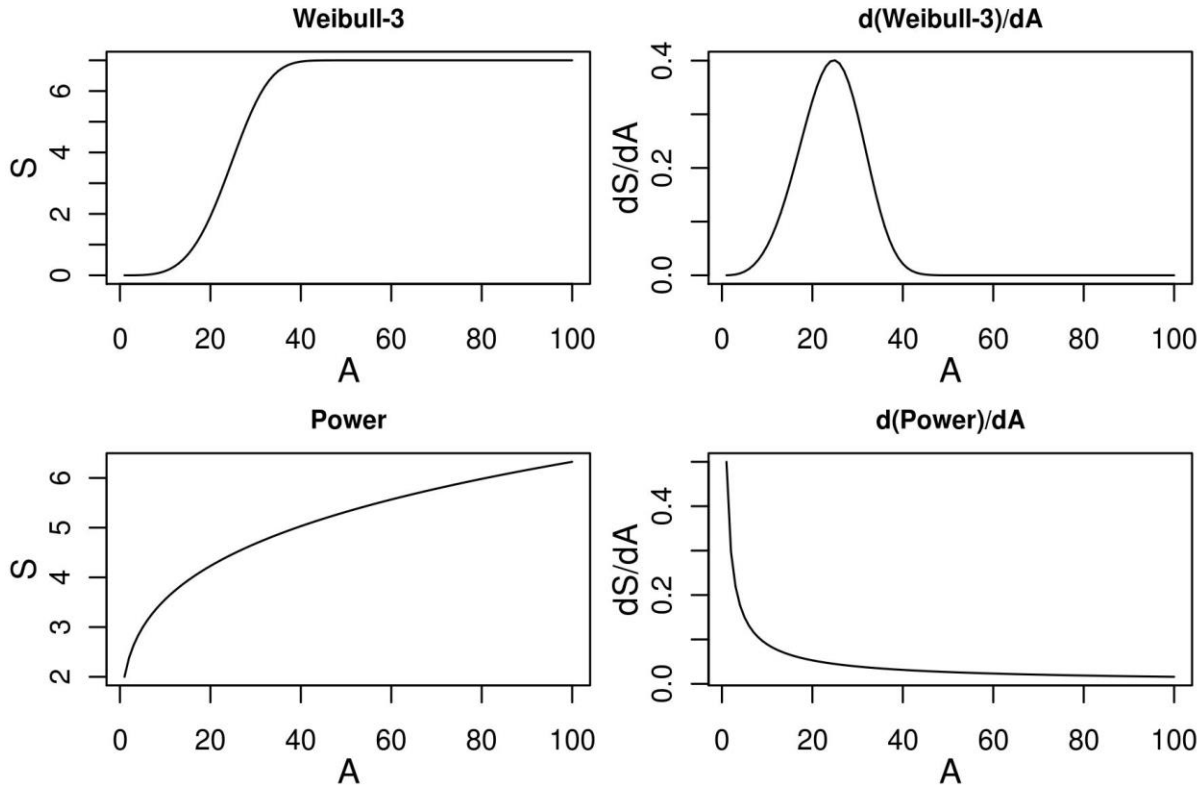


Figure A3.7 The functional form and first derivative of two ISAR models: the Weibull-3 and power models. The two left hand side plots show the functional form of the ISAR models: the plots of number of species (S) against island area (A). The right hand side plots illustrate the first derivative of the models, that is, the rate in change of species richness with increasing island area (dS/dA). The power model is a convex model, whilst the Weibull-3 model is sigmoidal (see Table 5.1 in the main paper).

Appendix 4

Supporting information: Island species–area relationships and species–accumulation curves are not equivalent data structures: a synthetic analysis of habitat island datasets (Chapter 6)

Appendix 4.1 The different types of species–area relationship

Whilst a number of different types of SAR have been formulated (e.g. Scheiner, 2003), in effect these different forms can be dichotomised into island species–area relationships ('ISARs') and species accumulation curves ('SACs'). The ISAR represents a curve whereby the number of species occurring within each of a set of islands is plotted as a function of the area of each island. The SAC refers to plots of increasing cumulative species number with increasing sampling effort/area. The SAC can be constructed in multiple ways that vary the order in which samples are accumulated to build the curve: in order of increasing or decreasing area or of increasing or decreasing species richness. Finally, the SAC can be constructed in a random manner, whereby the final curve represents an average of multiple permutations of random SAC construction. In our analysis, we construct the random SAC using a simple bootstrap procedure in which the data points were generated by randomly selecting the order of island addition into the SAC at each iteration, and taking the average data point values of 1000 iterations. While a variety of other methods have been proposed to calculate the random SAC (e.g. Colwell & Coddington, 1994; Ugland *et al.*, 2003), we used our bootstrap method as it was (a) relatively simple to compute, (b) enabled easy comparison and integration with our methods for calculating the other SAC structures, (c) has been shown to work in a similar context (Flather, 1996), and (d) does not require abundance data. The order in which samples are combined to construct the SAC determines the shape of the resultant curve, due to e.g. turnover of species between samples and sampling error (Gray *et al.*, 2004). These five types of SAC, in addition to the ISAR, are summarised in Table A4.1 below. Figure A4.1 illustrates how these different data structures can result in different curve shapes, when constructed from the same underlying data.

Colwell, R.K. & Coddington, J.A. (1994) Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **345**, 101-118.

Flather, C. (1996) Fitting species–accumulation functions and assessing regional land use impacts on avian diversity. *Journal of Biogeography*, **23**, 155-168.

Gray, J.S., Ugland, K.I. & Lambshead, J. (2004) Species accumulation and species–area curves — a comment on Scheiner (2003). *Global Ecology and Biogeography*, **13**, 473-476.

Scheiner, S.M. (2003) Six types of species–area curves. *Global Ecology and Biogeography*, **12**, 441-447.

Ugland, K.I., Gray, J.S. & Ellingsen, K.E. (2003) The species–accumulation curve and estimation of species richness. *Journal of Animal Ecology*, **72**, 888-897.

Table A4.1 The different types of species–area relationship data structures compared in this study. The acronyms are used throughout the paper. SAC refers to the species–accumulation curve.

Data structure	Acronym	Explanation
Island species–area relationship	ISAR	The number of species occurring within each of a set of islands is plotted as a function of the area of each isolate
Small–large SAC	SL	The order of island incorporation into the SAC starts with the smallest island and increases up to the largest
Large–small SAC	LS	The order of island incorporation into the SAC starts with the largest area and works down to the smallest island.
Poor–rich SAC	PR	The order of incorporation starts with the most species-poor island and, and increases up to the most species-rich island.
Rich–poor SAC	RP	The order of incorporation starts with the most species-rich island and up to the most species rich island.
Random SAC	Ran	The order of island accumulation is random.

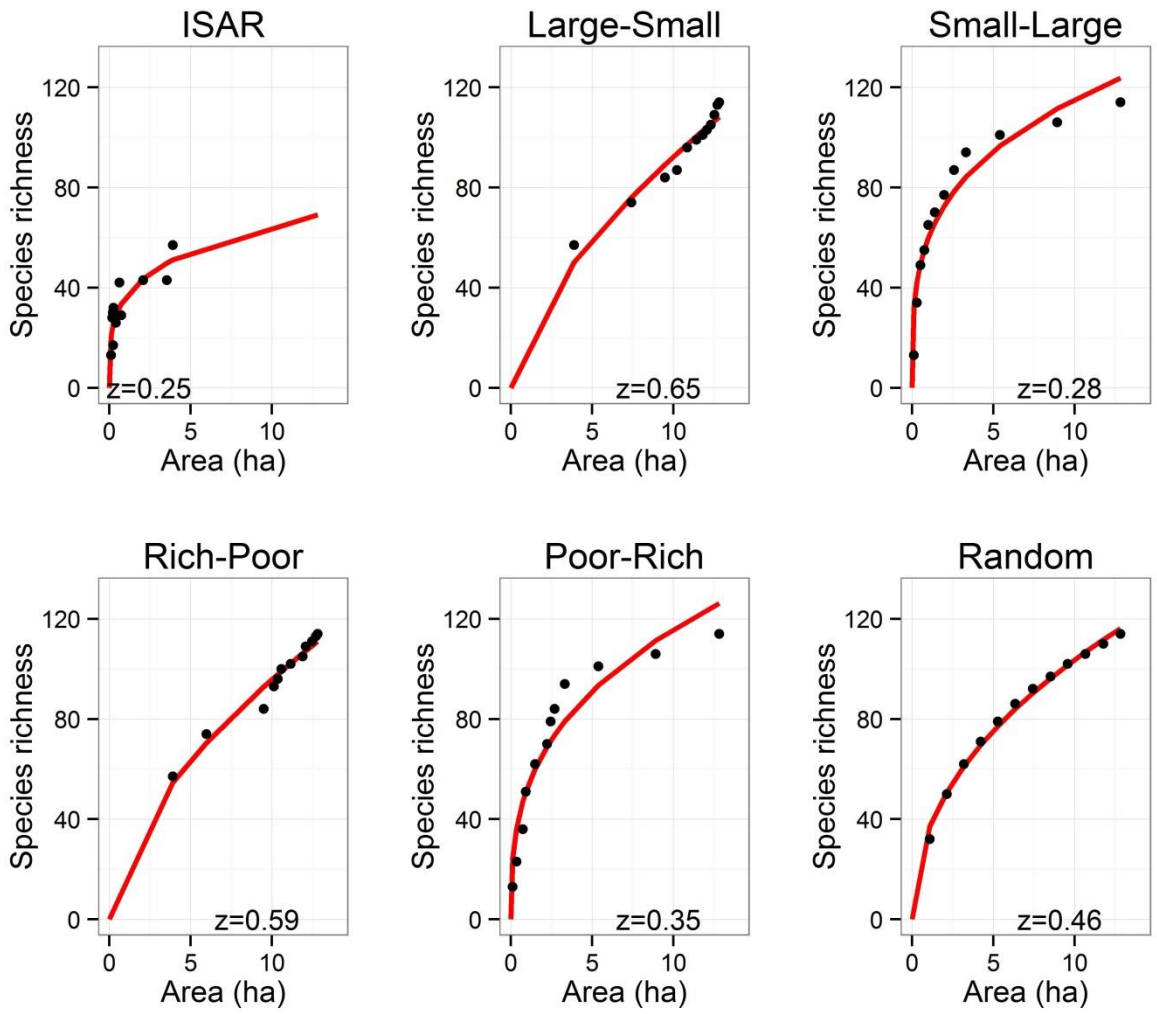


Figure A4.1 Exemplar fits of the power (non-linear) species–area relationship model to six different data structures, constructed from the same habitat island dataset: Gavish *et al.* (2012; invertebrates in a study system from Israel; number of islands = 12; number of species = 114). For each plot data for each of the individual habitat islands (black circles) are shown, alongside the fit of the power model (solid red line). The *z* value of the model is given in each instance. In the ISAR plot (top left), the fit of the power model has been extrapolated to the total cumulative area of all habitat islands in the system using the parameters derived from the model fitting process. That is, the extrapolated area data point was not included in the model fitting process. The curves have been forced through the origin.

Appendix A4.2 Source paper information

Classifying the matrix

For each dataset we recorded the predominant land-use type of the matrix surrounding the habitat islands using information in the source papers. Where possible we also checked this information using Google Earth, although this check was undertaken tentatively as many datasets were compiled decades ago and it is likely that the land-use has changed in the interim. These land-use categories were then converted into a 5-point scale of matrix permeability in which 1 represented the most permeable matrices, and 5 represented the less permeable matrices. This classification scale was determined using the results of Matthews *et al.* (*in review*), a study which examined how the slope of the island species–area relationship varied across multiple habitat island types. Obviously, this is a coarse approach, but obtaining fine scale matrix data for such a large number of studies is infeasible; particularly as such information is rarely, if ever, presented in the source papers. The classification was as follows:

1 – Secondary/re-growth forest (if the island type was forest) or scrub (if the island type was grassland)

2 –Scrub/grassland (if the island type was forest)

3 – Agriculture (e.g. crops, pasture)

4 – All mountaintop islands were included in this category as the matrix between them is known to be relatively impermeable (see Matthews *et al.*, *in review*).

5 – Urban land-uses and water (i.e. flooded forest fragment archipelagos)

Matthews, T.J., Guilhaumon, F., Triantis, K.A, Borregaard, M.K., & Whittaker, R.J. (*in review*) On the form of species–area relationships in habitat islands and true islands. *Global Ecology & Biogeography*.

Table A4.2 Description of the source papers, including the taxon, number of islands (No. Islands), the minimum island area in a dataset (Min. Area), and the number of species (No. Sp.). The ArcRes value for each dataset is also provided. The full references follow the table.

Source	Taxon	No. Islands	Min. Area	No. Sp.	ArcRes
Báldi & Kisbenedek (1999)	Invertebrates	27	0	36	0.2
Behle (1978)	Vertebrates	14	2849	75	0.01
Benedick *et al.* (2006)	Invertebrates	8	120.2	79	0.17
Blake & Karr 1984	Vertebrates	12	1.8	26	-0.1
Brotons & Herrando (2001)	Vertebrates	51	0.1	37	0.1
Brown (1971)	Vertebrates	17	3108	13	-0.19
Brown (1978)	Vertebrates	13	12173	11	0.06
Brown (1978)	Vertebrates	19	3108	16	-0.08
Cabrera-Guzmán & Reynoso (2012)	Vertebrates	6	1.4	35	-0.01
Castelletta *et al.* 2005	Vertebrates	6	7	132	0.05
Charles & Ang (2010)	Vertebrates	9	0	18	0.29
Cieślak & Dombrowski (1993)	Vertebrates	50	0	50	-0.41
Crooks (2002)	Vertebrates	29	2	7	0.18

Crowe (1979)	Plants	26	0	128	0.04
Daily & Ehrlich (1995)	Invertebrates	8	3	38	0.12
Dalecky *et al.* (2002)	Vertebrates	38	0.1	28	-0.31
Darlington *et al.* (2001)	Invertebrates	25	0	23	-0.14
Davies *et al.* (2003)	Invertebrates	10	0.2	95	0.2
Davis *et al.* (1988)	Vertebrates	12	800	12	-0.73
Dickman (1987)	Vertebrates	9	0.2	16	-0.02
Dickman (1987)	Vertebrates	12	0.2	20	0.13
Dinesen *et al.* (2001)	Vertebrates	18	100	9	0.19
dos Anjos & Boçon 1999	Vertebrates	12	0.5	188	0.03
dos Santos *et al.* (2007)	Plants	13	12.4	248	0.17
Edenius & Sjöberg (1997)	Vertebrates	18	2	23	-0.14
Essl & Dirnböck (2012)	Invertebrates	60	0	33	0.06
Feeley (2003)	Vertebrates	26	0.2	43	0.01
Fernández-Juricic (2000)	Vertebrates	25	1	32	0.08
Filgueiras *et al.* (2011)	Invertebrates	19	10.1	30	0.24
Flaspohler *et al.* 2010	Vertebrates	18	0.1	10	-0.02
Ford (1987)	Vertebrates	20	0.1	48	0
Galle (2008)	Invertebrates	15	0	51	0.25
Galli *et al.* (1976)	Vertebrates	10	0	35	-0.11
Ganzhorn *et al.* (1999)	Vertebrates	7	3	10	0.05
Gaublomme *et al.* (2008)	Invertebrates	10	5.3	91	0.21
Gavish *et al.* (2012)	Invertebrates	12	0.1	114	0.09
Gavish *et al.* (2012)	Invertebrates	12	0.1	115	0.1
Gillespie & Walter (2001)	Vertebrates	7	420	69	0.09
Haila *et al.* (1993)	Vertebrates	13	0.7	29	0.07
Hatt (1948)	Vertebrates	15	1	10	-0.44
Hatt (1948)	Vertebrates	15	1	152	-0.01
Hatt (1948)	Vertebrates	15	1	31	-0.34
Hattori & Ishida (2000)	Plants	38	0	177	0.01
Holbech (2005)	Vertebrates	15	2360	147	0.03
Hu *et al.* (2012)	Plants	152	0	383	0.2
Ishida *et al.* (1998)	Plants	29	0	119	0.02
Johnson (1975)	Vertebrates	20	3367	74	0.05
Kelt (2000)	Vertebrates	14	1.9	9	0.35
Kitchener *et al.* (1980)	Vertebrates	23	34	69	0.13
Kitchener *et al.* (1980)	Vertebrates	23	34	24	0.04
Kratter (1992)	Vertebrates	20	1	48	0.07
Langrand (1995)	Vertebrates	8	0.6	51	-0.04
Lomolino & Davis (1997)	Vertebrates	24	600	23	-0.07
Lomolino & Perault (2001)	Vertebrates	20	0.9	18	0.17
Lomolino *et al.* (1989)	Vertebrates	27	689	23	-0.08
Lumaret *et al.* (1997)	Plants	8	0	193	0.19

Maldonado-Coelho & Marini (2003)	Vertebrates	9	3.6	72	0.07
Marini (2001)	Vertebrates	6	7.5	145	0.06
Matthiae & Stearns (1981)	Vertebrates	22	0.4	13	-0.04
McCollin (1993)	Vertebrates	16	0.7	49	0.06
Meynard & Quinn (2008)	Vertebrates	9	180	51	0.14
Miyashita *et al.* (1998)	Invertebrates	7	0.2	34	0.04
Miyashita *et al.* (1998)	Invertebrates	9	0.4	52	0.06
Mohd-Azlan & Lawes (2011)	Vertebrates	13	0.7	70	0.1
Newmark (1991)	Vertebrates	10	0.1	31	-0.01
Nores (1995)	Vertebrates	9	1300	54	0.04
Nufio *et al.* (2011)	Invertebrates	13	0.8	38	0.04
Nyeko (2009)	Invertebrates	6	10	45	0.08
Peltzer *et al*. (2003)	Vertebrates	8	0	20	-0.02
Pineda & Halffter (2004)	Vertebrates	5	16	21	0.13
Ramanamanjato (2000)	Vertebrates	11	10	21	-0.34
Ramanamanjato (2000)	Vertebrates	11	10	41	-0.04
Ribas *et al.* (2005)	Invertebrates	18	3	136	0.01
Rosenblatt *et al.* (1999)	Vertebrates	10	1.8	16	0.02
Ruiz-Gutiérrez *et al.* (2008)	Vertebrates	7	1.5	184	0.04
Shreeve & Mason (1980)	Invertebrates	22	2	26	-0.28
Silva (2001)	Vertebrates	14	0.1	11	0.26
Silva & Pôrto (2009)	Plants	10	23	99	0.19
Simberloff & Martin (1991)	Vertebrates	35	0.4	45	-0.41
Smith *et al.* (1996)	Vertebrates	26	0.5	19	0.1
Suarez *et al.* (1998)	Invertebrates	40	0.4	45	0.1
Summerville *et al.* (2002)	Invertebrates	25	0	30	-0.59
Matthews *et al.* (2014)	Vertebrates	40	0.1	60	0.05
Matthews *et al.* (2014)	Vertebrates	42	0.1	47	0.08
Matthews *et al.* (2014)	Vertebrates	41	0.2	55	-0.05
Matthews *et al.* (2014)	Vertebrates	37	0.5	43	-0.05
Tonn & Magnuson (1982)	Vertebrates	18	2.4	23	0.06
Usher & Keiller (1998)	Invertebrates	18	0.5	214	0.12
Vallan (2000)	Vertebrates	7	0.2	28	0.05
Viveiros de Castro & Fernandez (2004)	Vertebrates	8	1.2	11	-0.04
Wang *et al.* (2010)	Vertebrates	14	1	11	0.23
Wang et al (2010)	Vertebrates	42	0.3	93	0.06
Watson (2003)	Vertebrates	17	2	129	0.05
Weaver & Kellman (1981)	Plants	10	0.8	30	0.15
Willson *et al.* (1994)	Vertebrates	10	0.5	25	0.06
Yong *et al.* (2011)	Vertebrates	6	1.1	75	0.06
Zimmerman & Bierregaard (1986)	Vertebrates	7	1	40	-0.03

Source paper references:

Báldi, A. & Kisbenedek, T. (1999) Orthopterans in small steppe patches: an investigation for the best-fit model of the species–area curve and evidences for their non-random distribution in the patches. *Acta Oecologica*, **20**, 125-132.

Behle, W.H. (1978) Avian biogeography of the Great Basin and intermountain region. *Great Basin Naturalist Memoirs*, **2**, 55-80.

Benedick, S., Hill, J.K., Mustaffa, N. et al. (2006) Impacts of rain forest fragmentation on butterflies in northern Borneo: species richness, turnover and the value of small fragments. *Journal of Applied Ecology*, **43**, 967-977.

Blake, J.G. & Karr, J.R. (1984) Species composition of bird communities and the conservation benefit of large versus small forests. *Biological Conservation*, **30**, 173-187.

Brotons, L. & Herrando, S. (2001) Reduced bird occurrence in pine forest fragments associated with road proximity in a Mediterranean agricultural area. *Landscape and Urban Planning*, **57**, 77-89.

Brown, J.H. (1971) Mammals on mountaintops: nonequilibrium insular biogeography. *The American Naturalist*, **105**, 467-478.

Brown, J.H. (1978) The theory of island biogeography and the distribution of boreal birds and mammals. *Great Basin Naturalist Memoirs*, **2,** 209-227.

Cabrera-Guzmán, E. & Reynoso, V. (2012) Amphibian and reptile communities of rainforest fragments: minimum patch size to support high richness and abundance. *Biodiversity and Conservation*, **21**, 3243-3265.

Castelletta, M., Thiollay, J.-M. & Sodhi, N.S. (2005) The effects of extreme forest fragmentation on the bird community of Singapore Island. *Biological Conservation*, **121**, 135-155.

Charles, J. & Ang, B. (2010) Non-volant small mammal community responses to fragmentation of kerangas forests in Brunei Darussalam. *Biodiversity and Conservation*, **19**, 543-561.

Cieślak, M. & Dombrowski, A. (1993) The effect of forest size on breeding bird communities. *Acta Ornithologica*, **27**, 97-111.

Crooks, K.R. (2002) Relative sensitivities of mammalian carnivores to habitat fragmentation. *Conservation Biology*, **16**, 488-502.

Crowe, T.M. (1979) Lots of weeds: insular phytogeography of vacant urban lots. *Journal of Biogeography*, **6**, 169-181.

Daily, G.C. & Ehrlich, P.R. (1995) Preservation of biodiversity in small rainforest patches: rapid evaluations using butterfly trapping. *Biodiversity and Conservation*, **4**, 35-55.

Dalecky, A., Chauvet, S., Ringuet, S. et al. (2002) Large mammals on small islands: short term effects of forest fragmentation on the large mammal fauna in French Guiana. *Revue d'écologie*, **57**, 145-164.

Darlington, J.P.E.C., Kaib, M. & Brand, R. (2001) Termites (Isoptera) in forest remnants and forest islands in the Shimba Hills National Reserve, coastal province of Kenya. *Sociobiology*, **37**, 527-538.

Davies, R.G., Hernández, L.M., Eggleton, P., Didham, R.K., Fagan, L.L. & Winchester, N.N. (2003) Environmental and spatial influences upon species composition of a termite assemblage across Neotropical forest islands. *Journal of Tropical Ecology***, 19**, 509-524.

Davis, R., Dunford, C. & Lomolino, M.V. (1988) Montane mammals of the American Southwest: the possible influence of post-Pleistocene colonization. *Journal of Biogeography*, **15**, 841-848.

Dickman, C.R. (1987) Habitat fragmentation and vertebrate species richness in an urban environment. *Journal of Applied Ecology*, **24**, 337-351.

Dinesen, L., Lehmberg, T., Rahner, M.C. & Fjeldså, J. (2001) Conservation priorities for the forests of the Udzungwa Mountains, Tanzania, based on primates, duikers and birds. *Biological Conservation*, **99**, 223-236.

dos Anjos, L. & Boçon, R. (1999) Bird communities in natural forest patches in southern Brazil. *The Wilson Bulletin*, **111**, 397-414.

dos Santos, K., Kinoshita, L.S. & Santos, F.A.M.d. (2007) Tree species composition and similarity in semideciduous forest fragments of southeastern Brazil. *Biological Conservation*, **135**, 268-277.

Edenius, L. & Sjöberg, K. (1997) Distribution of birds in natural landscape mosaics of old-growth forests in northern Sweden: relations to habitat area and landscape context. *Ecography*, **20**, 425-431.

Essl, F. & Dirnböck, T. (2012) What determines Orthoptera species distribution and richness in temperate semi-natural dry grassland remnants? *Biodiversity and Conservation*, **21**, 2525-2537.

Feeley, K. (2003) Analysis of avian communities in Lake Guri, Venezuela, using multiple assembly rule models. *Oecologia*, **137**, 104-113.

Fernández-Juricic, E. (2000) Bird community composition patterns in urban parks of Madrid: the role of age, size and isolation. *Ecological Research*, **15**, 373-383.

Filgueiras, B.K.C., Iannuzzi, L. & Leal, I.R. (2011) Habitat fragmentation alters the structure of dung beetle communities in the Atlantic Forest. *Biological Conservation*, **144**, 362-369.

Flaspohler, D.J., Giardina, C.P., Asner, G.P., Hart, P., Price, J., Lyons, C.K. & Castaneda, X. (2010) Long-term effects of fragmentation and fragment properties on bird species richness in Hawaiian forests. *Biological Conservation*, **143**, 280-288.

Ford, H.A. (1987) Bird communities on habitat islands in England. *Bird Study*, **34**, 205 - 218.

Galle, R. (2008) The effect of a naturally fragmented landscape on the spider assemblages. *North-Western Journal of Zoology*, **4**, 61-71.

Galli, A.E., Leck, C.F. & Forman, R.T.T. (1976) Avian distribution patterns in forest islands of different sizes in central New Jersey. *The Auk*, **93**, 356-364.

Ganzhorn, J.U., Fietz, J., Rakotovao, E., Schwab, D. & Zinner, D. (1999) Lemurs and the regeneration of dry deciduous forest in Madagascar. *Conservation Biology*, **13**, 794-804.

Gaublomme, E., Hendrickx, F., Dhuyvetter, H. & Desender, K. (2008) The effects of forest patch size and matrix type on changes in carabid beetle assemblages in an urbanized landscape. *Biological Conservation*, **141**, 2585-2596.

Gavish, Y., Ziv, Y. & Rosenzweig, M.L. (2012) Decoupling fragmentation from habitat loss for spiders in patchy agricultural landscapes. *Conservation Biology*, **26**, 150-159.

Gillespie, T.W. & Walter, H. (2001) Distribution of bird species richness at a regional scale in tropical dry forest of Central America. *Journal of Biogeography*, **28**, 651-662.

Haila, Y., Hanski, I.K. & Raivio, S. (1993) Turnover of breeding birds in small forest fragments: the "sampling" colonization hypothesis corroborated. *Ecology*, **74,** 714-725.

Hatt, R. (1948) *Island life: a study of land vertebrates of the islands of eastern Lake Michigan.* Cranbrook Institute of Science, Michigan.

Hattori, T. & Ishida, H. (2000) Relationship between species diversity, species composition and forest area of fragmented lucidophyllous forests in central Miyazaki Prefecture. *Japanese Journal of Ecology*, **50**, 221-234.

Holbech, L.H. (2005) The implications of selective logging and forest fragmentation for the conservation of avian diversity in evergreen forests of south-west Ghana. *Bird Conservation International*, **15**, 27-52.

Hu, G., Wu, J., Feeley, K.J., Xu, G. & Yu, M. (2012) The effects of landscape variables on the species-area relationship during late-stage habitat fragmentation. *PloS One*, **7**, e43894.

Ishida, H., Hattori, T., Takeda, Y. & Kodate, S. (1998) Relationship between species richness or species composition and area of fragmented lucidophyllous forests in southeastern Hyogo Prefecture. *Japanese Journal of Ecology*, **48,** 1-16.

Johnson, N.K. (1975) Controls of number of bird species on montane islands in the great basin. *Evolution*, **29**, 545-567.

Kelt, D.A. (2000) Small mammal communities in rainforest fragments in central southern Chile. *Biological Conservation*, **92**, 345-358.

Kitchener, D.J., Chapman, A., Dell, J., Muir, B.G. & Palmer, M. (1980a) Lizard assemblage and reserve size and structure in the Western Australian wheatbelt — some implications for conservation. *Biological Conservation*, **17**, 25-62.

Kitchener, D.J., Chapman, A., Muir, B.G. & Palmer, M. (1980b) The conservation value for mammals of reserves in the western Australian wheatbelt. *Biological Conservation*, **18**, 179-207.

Kratter, A.W. (1992) Montane avian biogeography in southern California and Baja California. *Journal of Biogeography*, **19,** 269-283.

Langrand, O. (1995) *The effects of forest fragmentation on bird species in Madagascar: a case study from Ambohitantely Forest Reserve on the central high plateau.* University of Natal, Pietermaritzburg.

Lomolino, M.V. & Davis, R. (1997) Biogeographic scale and biodiversity of mountain forest mammals of western North America. *Global Ecology and Biogeography Letters*, **6,** 57-76.

Lomolino, M.V. & Perault, D.R. (2001) Island biogeography and landscape ecology of mammals inhabiting fragmented, temperate rain forests. *Global Ecology and Biogeography*, **10**, 113-132.

Lomolino, M.V., Brown, J.H. & Davis, R. (1989) Island biogeography of montane forest mammals in the American Southwest. *Ecology*, **70**, 180-194.

Lumaret, R., Guillerm, J.-L., Maillet, J. & Verlaque, R. (1997) Plant species diversity and polyploidy in islands of natural vegetation isolated in extensive cultivated lands. *Biodiversity and Conservation*, **6**, 591-613.

Maldonado-Coelho, M. & Marini, M.Â. (2000) Effects of forest fragment size and successional stage on mixed-species bird flocks in southeastern Brazil. *The Condor*, **102**, 585-594.

Marini, M.Â. (2001) Effects of forest fragmentation on birds of the Cerrado region, Brazil. *Bird Conservation International*, **11,** 13-25.

Matthews, T.J., Cottee-Jones, H.E. & Whittaker, R.J. (2014) Habitat fragmentation and the species–area relationship: a focus on total species richness obscures the impact of habitat loss on habitat specialists. *Diversity and Distributions*, **20**, 1136-1146.

Matthiae, P.E. & Stearns, F. (1981) Mammals in forests islands in southeastern Wisconsin. *Forest Island Dynamics in Man-Dominated Landscapes.* (ed. by R.L. Burgess and D.M. Sharpe), pp. 55-66. Springer-Verlag, New York.

McCollin, D. (1993) Avian distribution patterns in a fragmented wooded landscape (north Humberside, U.K.): the role of between-patch and within-patch structure. *Global Ecology and Biogeography Letters*, **3**, 48-62.

Meynard, C.N. & Quinn, J.F. (2008) Bird metacommunities in temperate South American forest: vegetation structure, area, and climate effects. *Ecology*, **89**, 981-990.

Miyashita, T., Shinkai, A. & Chida, T. (1998) The effects of forest fragmentation on web spider communities in urban areas. *Biological Conservation***, 86,** 357-364.

Mohd-Azlan, J. & Lawes, M.J. (2011) The effect of the surrounding landscape matrix on mangrove bird community assembly in north Australia. *Biological Conservation*, **144**, 2134-2141.

Newmark, W.D. (1991) Tropical forest fragmentation and the local extinction of understory birds in the eastern Usambara Mountains, Tanzania. *Conservation Biology*, **5**, 67-78.

Nores, M. (1995) Insular biogeography of birds on mountain-tops in north western Argentina. *Journal of Biogeography*, **22**, 61-70.
Nufio, C., McClenahan, J. & Deane Bowers, M. (2011) Grasshopper response to reductions in habitat area as mediated by subfamily classification and life history traits. *Journal of Insect Conservation*, **15**, 409-419.
Nyeko, P. (2009) Dung beetle assemblages and seasonality in primary forest and forest fragments on agricultural landscapes in Budongo, Uganda. *Biotropica*, **41**, 476-484.
Peltzer, P.M., Lajmanovich, R.C. & Beltzer, A.H. (2003) The effects of habitat fragmentation on amphibian species richness in the floodplain of the Middle Parana River, Argentina. *Herpetological Journal*, **13**, 95-98.
Pineda, E. & Halffter, G. (2004) Species diversity and habitat fragmentation: frogs in a tropical montane landscape in Mexico. *Biological Conservation*, **117**, 499-508.
Ramanamanjato, J.-B. (2000) Fragmentation effects on reptile and amphibian diversity in the littoral forest of southeastern Madagascar. *Isolated Vertebrate Communities in the Tropics.* (ed. by G. Rheinwald), pp. 297-308. Zoologisches Forschungsinstitut und Museum A. Koenig, Bonn.
Ribas, C.R., Sobrinho, T.G., Schoereder, J.H., Sperber, C.F., Lopes-Andrade, C. & Soares, S.M. (2005) How large is large enough for insects? Forest fragmentation effects at three spatial scales. *Acta Oecologica*, **27**, 31-41.
Rosenblatt, D.L., Heske, E.J., Nelson, S.L., Barber, D.M., Miller, M.A. & MacAllister, B. (1999) Forest fragments in east-central Illinois: islands or habitat patches for mammals? *The American Midland Naturalist*, **141**, 115-123.
Ruiz-Gutiérrez, V., Gavin, T.A. & Dhondt, A.A. (2008) Habitat fragmentation lowers survival of a tropical forest bird. *Ecological Applications*, **18**, 838-846.
Shreeve, T.G. & Mason, C.F. (1980) The number of butterfly species in woodlands. *Oecologia*, **45**, 414-418.
Silva, M. (2001) Abundance, diversity, and community structure of small mammals in forest fragments in Prince Edward Island National Park, Canada. *Canadian Journal of Zoology*, **79**, 2063-2071.
Silva, M. & Pôrto, K. (2009) Effect of fragmentation on the community structure of epixylic bryophytes in Atlantic Forest remnants in the northeast of Brazil. *Biodiversity and Conservation*, **18**, 317-337.
Simberloff, D. & Martin, J.L. (1991) Nestedness of insular avifaunas: simple summary statistics masking complex species patterns. *Ornis Fennica*, **68**, 178-192.
Smith, G.T., Arnold, G.W., Sarre, S., Abensperg-Traun, M. & Steven, D.E. (1996) The effect of habitat fragmentation and livestock grazing on animal communities in remnants of gimlet eucalyptus salubris woodland in the western Australian wheatbelt. II. Lizards. *Journal of Applied Ecology*, **33**, 1302-1310.
Suarez, A.V., Bolger, D.T. & Case, T.J. (1998) Effects of fragmentation and invasion on native ant communities in coastal southern California. *Ecology*, **79**, 2041-2056.
Summerville, K.S., Veech, J.A. & Crist, T.O. (2002) Does variation in patch use among butterfly species contribute to nestedness at fine spatial scales? *Oikos*, **97**, 195-204.
Tonn, W.M. & Magnuson, J.J. (1982) Patterns in the species composition and richness of fish assemblages in northern Wisconsin lakes. *Ecology*, **63**, 1149-1166.
Usher, M.B. & Keiller, S.W.J. (1998) The macrolepidoptera of farm woodlands: determinants of diversity and community structure. *Biodiversity and Conservation*, **7**, 725-748.
Vallan, D. (2000) Influence of forest fragmentation on amphibian diversity in the nature reserve of Ambohitantely, highland Madagascar. *Biological Conservation*, **96**, 31-43.
Viveiros de Castro, E.B. & Fernandez, F.A.S. (2004) Determinants of differential extinction vulnerabilities of small mammals in Atlantic forest fragments in Brazil. *Biological Conservation*, **119**, 73-80.

Wang, Y., Bao, Y., Yu, M., Xu, G. & Ding, P. (2010) Nestedness for different reasons: the distributions of birds, lizards and small mammals on islands of an inundated lake. *Diversity and Distributions*, **16**, 862-873.
Watson, D.M. (2003) Long-term consequences of habitat fragmentation — highland birds in Oaxaca, Mexico. *Biological Conservation*, **111**, 283-303.
Weaver, M. & Kellman, M. (1981) The effects of forest fragmentation on woodlot tree biotas in southern Ontario. *Journal of Biogeography*, **8**, 199-210.
Willson, M.F., De Santo, T.L., Sabag, C. & Armesto, J.J. (1994) Avian communities of fragmented south-temperate rainforests in Chile. *Conservation Biology*, **8**, 508-520.
Yong, D.L., Qie, L., Sodhi, N.S., Koh, L.P., Peh, K.S.-H., Lee, T.M., Lim, H.C. & Lim, S.L.-H. (2011) Do insectivorous bird communities decline on land-bridge forest islands in Peninsular Malaysia? *Journal of Tropical Ecology*, **27**, 1-14.
Zimmerman, B.L. & Bierregaard, R.O. (1986) Relevance of the equilibrium–theory of island biogeography and species area relations to conservation with a case from Amazonia. *Journal of Biogeography*, **13**, 133-143.

Appendix A4.3 Expanded study methods

SAR model comparison

We fit a set of 20 SAR models to both the ISAR and SAC forms of each dataset. After unsatisfactory fits (e.g. fits with non-normal residuals, or fits in which in the model fitting algorithm did not converge) were removed, we considered the model with the lowest Akaike's information criterion corrected for small sample size (AIC_c), as providing the best fit (Burnham & Anderson, 2002). The 20 models are listed as having specific shapes (i.e. either of linear, convex or sigmoidal). However, the actual observed shape of the model fits can vary from these listed shapes. For example, sigmoidal models can produce observed convex or linear fits depending on the parameter values. Thus, we used the algorithm of Triantis *et al.* (2012) to determine the observed shape of the best fitting model. This algorithm works by first determining whether the observed shape of a model fit was linear. This is achieved by comparing the shape of the model fit with a straight line which joins the fitted values for the smallest and largest island area in the dataset. One hundred equally spaced increments between the smallest and largest island area are created, and the shape of the fit is classified as linear if all of the differences between the number of species calculated from the straight line and from the model fit are less than 0.001 species. If the shape of the fit was not classified as linear, the second derivative of the model fit was examined for an inflexion point to distinguish between a convex and sigmoid shape. If the second derivative changed sign, the model fit displayed an inflection point in the observed range of the data, and thus the fit was classified as sigmoid; if not, convex.

Determining the degree of deviation of the archipelagic point

We plotted the ISAR data in log–log space and fitted the power (log–log; base 10) model to these data. We then calculated the archipelagic point as the cumulative total area of all habitat islands in the dataset plotted against the cumulative species richness total; and predicted the number of species in the archipelagic point ('SApred') using the ISAR model of the constituent islands. To determine if SAobs deviated from SApred we used one of the simple indices employed by Santos *et al.* (2010), first calculating the residuals of the ISAR model (observed minus predicted species richness) and then selecting the maximum absolute value (i.e. MaxRes). MaxRes was then expressed as a proportion of the predicted log(species richness) of the island from which MaxRes was generated from (PropMaxRes). SAobs was said to deviate from SApred if SAobs was greater than SApred ± (SApred x PropMaxRes).

We also followed Santos *et al.* (2010) and calculated the archipelagic residual ('ArcRes') as the difference between log(SAobs) and SApred, standardised by log(SAobs).

Appendix A4.4 Supplementary results

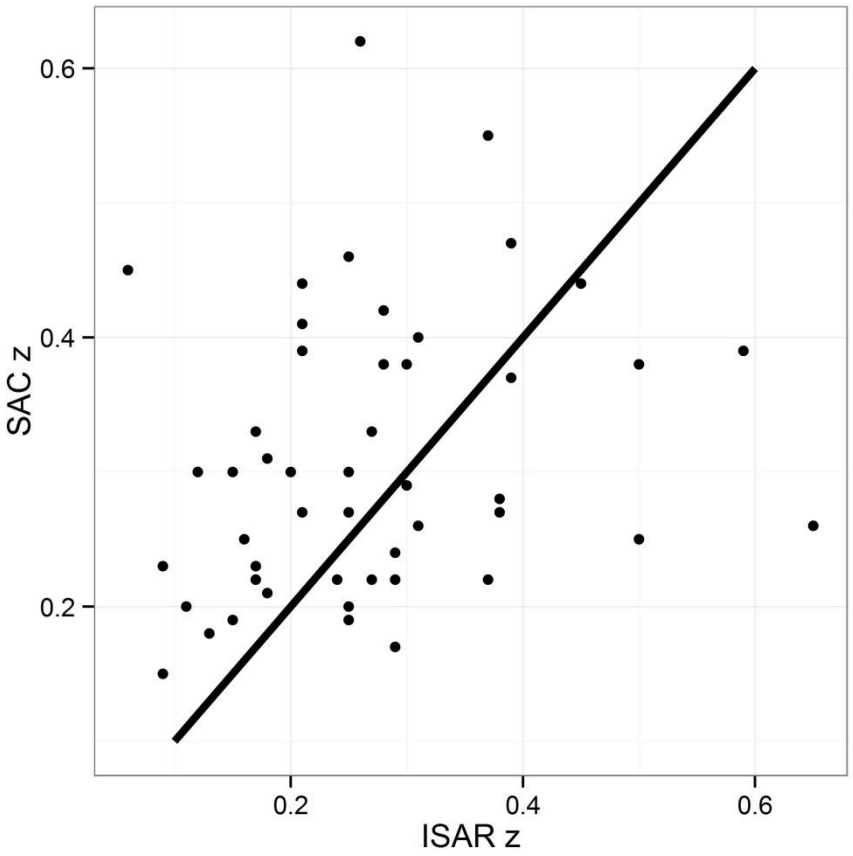


Figure A4.2 The *z* value of the power model fitted using ISAR structured data plotted against *z* values derived using randomly constructed SAC data (see Table A4.1), for 50 habitat island datasets. The solid line represents a 1:1 fit line; points below the line represent datasets in which the ISAR z value was larger than the SAC *z* value, and vice versa.

Appendix 5

Supporting information: Habitat fragmentation and the species–area relationship: a focus on total species richness obscures the impact of habitat loss on habitat specialists (Chapter 7)

Appendix A5.1 Source paper information and species classification methodology

Table A5.1 Description of the source papers. The habitat island type in each case was forest. The full references follow the table.

Study	Location	No. of hab. isl.	No. of sp.	Area range (Ha)	Richness range	Matrix type
1. Anciães & Marini (2000)	Brazil	7	100	15-350	17-25	Agriculture
2. Blake & Karr(1984)	USA	12	26	1.8-600	4-24	Agriculture
3. Castelletta *et al.* (2005)	Singapore	10	132	7-935	49-98	Urban & Grassland
4. Cieślak & Dombrowski (1993)	Poland	50	50	0.04-15.44	0-37	Agriculture
5. dos Anjos & Boçon (1999)	Brazil	12	119	0.5-10	45-92	Grassland
6. Ford (1987)	UK	20	48	0.14-18	13-35	Agriculture
7. Forman (1976)	USA	10	31	0.01-24	0-31	Agriculture
8. Gillespie & Walter (2001)	Costa Rica	7	69	420-16804	20-45	Agriculture
9. Hannah *et al.*(2007)	Australia	13	127	0.0002-0.2044	24-73	Agriculture
10. Holbech (1995)	Ghana	15	147	2360-58790	80-117	Savannah
11. Langrand (1995)	Madagascar	8	51	0.64-1250	15-51	Grassland
12. Marini (2001)	Brazil	6	145	7.5-230	46-103	Agriculture
13. This study	France	40	47	0.11-150	5-40	Agriculture
14. This study	Norway	41	55	0.2-150	2-32	Agriculture
15. This study	Spain	42	47	0.06-110	6-34	Agriculture
16. This study	UK	37	43	0.5-172	4-32	Agriculture
17. McCollin (1993)	UK	16	49	0.74-14.51	12-32	Agriculture
18. Ríos *et al.* (2011)	Peru	5	52	4.3-82	15-28	Grassland
19. Simberloff & Martin (1991)	Finland	35	45	0.4-101	0-30	Agriculture
20. Watson (2003)	Mexico	17	129	2-159242	4-78	Scrub
21. Wethered & Lawes (2005) - Balgowan	South Africa	9	55	0.5-215	23-49	Grassland & Plantation
22. Wethered & Lawes (2005) - Gilgoa	South Africa	9	52	0.5-273	28-39	Grassland & Plantation
23. Willson *et al.* (1994)	Chile	10	25	0.5-350	8-20	Agriculture

Source papers:

Anciães, M. & Marini, M.Â. (2000) The effects of fragmentation on fluctuating asymmetry in passerine birds of Brazilian tropical forests. *Journal of Applied Ecology*, **37**, 1013–1028.

Blake, J.G. & Karr, J.R. (1984) Species composition of bird communities and the conservation benefit of large versus small forests. *Biological Conservation*, **30**, 173–187.

Castelletta, M., Thiollay, J.-M. & Sodhi, N.S. (2005) The effects of extreme forest fragmentation on the bird community of Singapore Island. *Biological Conservation*, **121**, 135–155.

Cieślak, M. & Dombrowski, A. (1993) The effect of forest size on breeding bird communities. *Acta Ornithologica*, **27**, 97–111.

dos Anjos, L. & Boçon, R. (1999) Bird communities in natural forest patches in Southern Brazil. *The Wilson Bulletin*, **111**, 397–414.

Ford, H.A. (1987) Bird communities on habitat islands in England. *Bird Study*, **34**, 205–218.

Forman, R.T.T., Galli, A.E. & Leck, C.F. (1976) Forest size and avian diversity in New Jersey woodlots with some land use implications. *Oecologia*, **26**, 1–8.

Gillespie, T.W. & Walter, H. (2001) Distribution of bird species richness at a regional scale in tropical dry forest of Central America. *Journal of Biogeography*, **28**, 651–662.

Hannah, D., Woinarski, J.C.Z., Catterall, C.P., McCosker, J.C., Thurgate, N.Y. & Fensham, R.J. (2007) Impacts of clearing, fragmentation and disturbance on the bird fauna of Eucalypt savanna woodlands in central Queensland, Australia. *Austral Ecology*, **32**, 261–276.

Holbech, L.H. (2005) The implications of selective logging and forest fragmentation for the conservation of avian diversity in evergreen forests of south-west Ghana. *Bird Conservation International*, **15**, 27–52.

Langrand, O. (1995) *The effects of forest fragmentation on bird species in Madagascar: a case study from Ambohitantely Forest Reserve on the central high plateau*. University of Natal, Pietermaritzburg.

Marini, M.Â. (2001) Effects of forest fragmentation on birds of the cerrado region, Brazil. *Bird Conservation International*, **11,** 13–25.

McCollin, D. (1993) Avian distribution patterns in a fragmented wooded landscape (North Humberside, U.K.): the role of between-patch and within-patch structure. *Global Ecology and Biogeography Letters*, **3**, 48–62.

Ríos, S.S., Lloyd, H. & Valdés-Velásquez, A. (2011) Bird species richness, diversity and abundance in Polylepis woodlands, Huascaran biosphere reserve, Peru. *Studies on Neotropical Fauna and Environment*, **46**, 69–76.

Simberloff, D. & Martin, J.L. (1991) Nestedness of insular avifaunas: simple summary statistics masking complex species patterns. *Ornis Fennica*, **68,** 178–192.

Watson, D.M. (2003) Long-term consequences of habitat fragmentation-highland birds in Oaxaca, Mexico. *Biological Conservation*, **111**, 283–303.

Wethered, R. & Lawes, M.J. (2005) Nestedness of bird assemblages in fragmented Afromontane forest: the effect of plantation forestry in the matrix. *Biological Conservation*, **123**, 125–137.

Willson, M.F., De Santo, T.L., Sabag, C. & Armesto, J.J. (1994) Avian communities of fragmented south-temperate rainforests in Chile. *Conservation Biology*, **8**, 508–520.

Species classification method and references

To classify species into specialists and generalists we used a combination of books, articles, online reference databases and expert opinion. The books were largely accessed through the

Alexander Library of Ornithology, University of Oxford. Where necessary, articles were obtained through a literature search using combinations of the following keywords: “birds”, “habitat specialisation”, “forest specialists”, “habitat specificity” and a given region, for example “South America”. A number of studies (e.g. Anciães & Marini, 2000; Salisbury *et al.*, 2012) had already classified a large number of species based on habitat specificity (e.g. Salisbury *et al.*, 2012, classified 739 species in the Amazon), which provided us with a resource with which to compare with our own classifications. Finally, we also consulted a number of ornithological experts for additional opinions on particular species.

The following references relate to the core source material used to classify the species in our datasets. A large number of additional books, articles (found using the aforementioned keywords and from the reference lists of sourced papers) and other sources of identification material from the Alexander Library and the wider literature were referenced for individual species information and are not included here:

Alonso, J.A. & Whitney, B.M. (2003) New distributional records of birds from white-sand forests of the northern Peruvian Amazon, with implications for biogeography of northern South America. *The Condor*, **105,** 552–566.

Anciães, M. & Marini, M.Â. (2000) The effects of fragmentation on fluctuating asymmetry in passerine birds of Brazilian tropical forests. *Journal of Applied Ecology*, **37**, 1013–1028.

Birdlife (2011) *Birdlife data zone*. Available at: http://www.birdlife.org/datazone/home.

Borges, S.H., Cohn-Haft, M., Carvalhaes, A.M.P., Henriques, L.M., Pacheco, J.F. & Whittaker, A. (2001) Birds of Jaú national park, Brazilian Amazon: Species check-list, biogeography and conservation. *Ornitologia Neotropical*, **12**, 109–140.

Borrow, N. & Demey, D. (2010) *Birds of Ghana*. Christopher Helm, London.

Brooker, M.G., Braithwaite, R.W. & Estbergs, J.A. (1990) Foraging ecology of some insectivorous and nectarivorous species of birds in forests and woodlands of the wet-dry tropics of Australia. *Emu*, **90**, 215–230.

Christiansen, M.B. & Pitter, E. (1997) Species loss in a forest bird community near Lagoa Santa in South Eastern Brazil. *Biological Conservation*, **80**, 23–32.

Cohn-Haft, M., Whittaker, A. & Stouffer, P.C. (1997) A new look at the "species-poor" Central Amazon: the avifauna north of Manaus, Brazil. *Ornithological Monographs*, **48**, 205–235.

Cramp, S. (ed.) (1980) *Handbook of the birds of Europe, the Middle East, and North Africa: the birds of the Western Palearctic. Vol. 2, Hawks to bustards.* Oxford University Press, Oxford.

Cramp, S. (ed.) (1985) *Handbook of the birds of Europe, the Middle East, and North Africa: the birds of the Western Palearctic*. *Vol. 4 Terns to woodpeckers*. Oxford University Press, Oxford.

Cramp, S. (ed.) (1988) *Handbook of the birds of Europe, the Middle East, and North Africa: the birds of the Western Palearctic*. *Vol. 5 Tyrant flycatchers to thrushes*. Oxford University Press, Oxford.

Cramp, S. & Brooks, D.J. (eds.) (1992) *Handbook of the birds of Europe, the Middle East, and North Africa: the birds of the Western Palearctic*. *Vol. 6 Warblers.* Oxford University Press, Oxford.

Cramp, S. & Perrins, C.M. (eds.) (1993) *Handbook of the birds of Europe, the Middle East, and North Africa: the birds of the Western Palearctic. Vol. 7 Flycatchers to shrikes.* Oxford University Press, Oxford.

Cramp, S. & Perrins, C.M. (eds.) (1994a) *Handbook of the birds of Europe, the Middle East, and North Africa: the birds of the western Palearctic. Vol. 8, Crows to finches*. Oxford University Press, Oxford.

Cramp, S. & Perrins, C.M. (eds.) (1994b) *Handbook of the birds of Europe, the Middle East and North Africa: the birds of the Western Palearctic. Vol. 9, Buntings and New World warblers*. Oxford University Press, Oxford.
del Hoyo, J. Elliott, A. & Sargatal, J. (eds.) (1992) *Handbook of the birds of the world. Vol. 1. Ostrich to ducks*. Lynx Edicions, Barcelona.
del Hoyo, J. Elliott, A. & Sargatal, J. (eds.) (1994) *Handbook of the birds of the world. Vol. 2. New World vultures to guineafowl*. Lynx Edicions, Barcelona.
del Hoyo, J. Elliott, A. & Sargatal, J. (eds.) (1996) *Handbook of the birds of the world. Vol. 3. Hoatzin to auks*. Lynx Edicions, Barcelona.
del Hoyo, J. Elliott, A. & Sargatal, J. (eds.) (1997) *Handbook of the birds of the world. Vol. 4. Sandgrouse to cuckoos*. Lynx Edicions, Barcelona.
del Hoyo, J. Elliott, A. & Sargatal, J. (eds.) (1999) *Handbook of the birds of the world. Vol. 5. Barn-owls to hummingbirds*. Lynx Edicions, Barcelona.
del Hoyo, J. Elliott, A. & Sargatal, J. (eds.) (2001) *Handbook of the birds of the world. Vol. 6. Mousebirds to hornbills*. Lynx Edicions, Barcelona.
del Hoyo, J. Elliott, A. & Sargatal, J. (eds.) (2002) *Handbook of the birds of the world. Vol. 7. Jacamars to woodpeckers*. Lynx Edicions, Barcelona.
del Hoyo, J. Elliott, A. & Christie, D.A. (eds.) (2003) *Handbook of the birds of the world. Vol. 8. Broadbills to tapaculos*. Lynx Edicions, Barcelona.
del Hoyo, J. Elliott, A. & Christie, D.A. (eds.) (2004) *Handbook of the birds of the world. Vol. 9. Cotingas to pipits and wagtails*. Lynx Edicions, Barcelona.
del Hoyo, J. Elliott, A. & Christie, D.A. (eds.) (2005) *Handbook of the birds of the world. Vol. 10. Cuckoo-shrikes to thrushes*. Lynx Edicions, Barcelona.
del Hoyo, J. Elliott, A. & Christie, D.A. (eds.) (2006) *Handbook of the birds of the world. Vol. 11. Old World flycatchers to Old World warblers*. Lynx Edicions, Barcelona.
del Hoyo, J. Elliott, A. & Christie, D.A. (eds.) (2007) *Handbook of the birds of the world. Vol. 12. Picathartes to tits and chickadees*. Lynx Edicions, Barcelona.
del Hoyo, J. Elliott, A. & Christie, D.A. (eds.) (2008) *Handbook of the birds of the world. Vol. 13. Penduline-tits to shrikes*. Lynx Edicions, Barcelona.
del Hoyo, J. Elliott, A. & Christie, D.A. eds. (2009) *Handbook of the birds of the world. Vol. 14. Bush-shrikes to Old World sparrows*. Lynx Edicions, Barcelona.
del Hoyo, J. Elliott, A. & Christie, D.A. (eds.) (2010) *Handbook of the birds of the world. Vol. 15. Weavers to New World warblers*. Lynx Edicions, Barcelona.
del Hoyo, J. Elliott, A. & Christie, D.A. (eds.) (2011) *Handbook of the birds of the world. Vol. 16. Tanagers to New World blackbirds*. Lynx Edicions, Barcelona.
Dorst, J. (1972) The evolution and affinities of the birds of Madagascar. *Biogeography and ecology in Madagascar* (ed. by R. Battistini and G. Richard-Vindard), pp. 615–627. Springer, Netherlands.
dos Anjos, L. (2001) Bird communities in five Atlantic forest fragments in southern Brazil. *Ornitologia Neotropical*, **12**, 11–27.
dos Anjos, L. & Boçon, R. (1999) Bird communities in natural forest patches in Southern Brazil. *The Wilson Bulletin*, **111**, 397–414.
Dunn, J. & Garrett, K. (1997) *A field guide to warblers of North America*. Houghton Mifflin, Boston.
Edwards, E.P. (1998) *A field guide to the birds of México and adjacent Areas: Belize, Guatemala, and El Salvador*. University of Texas Press, Austin.
Erize, F., Rodriguez Mata, J.R. & Rumboll, M. (2006) *Birds of South America: Non-Passerines: rheas to woodpeckers*. Princeton University Press, Princeton.
Finlayson, C. & Tomlinson, D. (2003) *Birds of Iberia.* Santana Books, Spain.
Ford, H.A. (1987) Bird communities on habitat islands in England. *Bird Study*, **34,** 205–218.
Garrigues, R. & Dean, R. (2007) *The birds of Costa Rica*. Cornell University Press, New York.

Grimmett, R., Inskipp, C. & Inskipp, T. (1999) *Guide to the birds of India, Pakistan, Nepal, Bangladesh, Bhutan, Sri Lanka, and the Maldives*. Princeton University Press, Princeton.

Griscom, L. & Greenway, J.C. (1941) Birds of lower Amazonia. *Bulletin of the Museum of Comparative Zoology*, **88**, 283–344.

Hawkins, A.F.A. (1999) Altitudinal and latitudinal distribution of east Malagasy forest bird communities. *Journal of Biogeography*, **26**, 447–458.

Holbech, L.H. (2005) The implications of selective logging and forest fragmentation for the conservation of avian diversity in evergreen forests of south-west Ghana. *Bird Conservation International*, **15**, 27–52.

Howell, S.N.G. & Webb, S. (1995) *A guide to the birds of Mexico and northern central America.* Oxford, Oxford University Press.

Jaramillo, A. (2003) *Birds of Chile.* Princeton University Press, Princeton.

Jeyarajasingam, A. & Pearson, A. (1999) *A field guide to the birds of west Malaysia and Singapore.* Oxford University Press, Oxford.

Krügel, M.M. & dos Anjos, L. (2000) Bird communities in forest remnants in the city of Maringá, Paraná State, Southern Brazil. *Ornitologia Neotropical*, **11**, 315–330.

Langrand, O. (1990) *Guide to the birds of Madagascar*. Yale University Press, New Haven.

Langrand, O. (1995) *The effects of forest fragmentation on bird species in Madagascar: a case study from Ambohitantely Forest Reserve on the central high plateau.* University of Natal, Pietermaritzburg.

Mac Nally, R., Bennett, A.F. & Horrocks, G. (2000) Forecasting the impacts of habitat fragmentation: evaluation of species-specific predictions of the impact of habitat fragmentation on birds in the box–ironbark forests of central Victoria, Australia. *Biological Conservation*, **95**, 7–29.

Marini, M.Â. (2001) Effects of forest fragmentation on birds of the Cerrado region, Brazil. *Bird Conservation International*, **11**, 13–25.

Martin, E.A., Viano, M., Ratsimisetra, L., Laloë, F. & Carrière, S.M. (2012) Maintenance of bird functional diversity in a traditional agroecosystem of Madagascar. *Agriculture, Ecosystems & Environment*, **149**, 1–9.

Naka, L.N., Cohn-Haft, M., Whittaker, A., Barnett, J.M. & Torres, M.D.E.F. (2007) Avian biogeography of Amazonian flooded forests in the Rio Branco Basin, Brazil. *The Wilson Journal of Ornithology*, **119**, 439–449.

Peh, K.S.H., Jong, J.d., Sodhi, N.S., Lim, S.L.H. & Yap, C.A.M. (2005) Lowland rainforest avifauna and human disturbance: persistence of primary forest birds in selectively logged forests and mixed-rural habitats of southern Peninsular Malaysia. *Biological Conservation*, **123**, 489–505.

Restall, R., Rodner, C. & Lentino, M. (2006a) *Birds of Northern South America: an identification guide. Vol 1*. Christopher Helm, London.

Restall, R., Rodner, C. & Lentino, M. (2006b) *Birds of Northern South America: an identification guide. Vol 2*. Christopher Helm, London.

Ridgely, R.S. & Greenfield, P.J. (2001a) *The birds of Ecuador, vol. I. Status, distribution and taxonomy*. Cornell University Press, New York.

Ridgely, R.S. & Greenfield, P.J. (2001b) *The birds of Ecuador, vol. II. Field guide.* Cornell University Press, New York.

Ridgely, R.S. & Tudor, G. (2009) *Field guide to the songbirds of South America: The passerines*. University of Texas Press, Austin.

Robson, C. (2000) *A field guide to the birds of South-East Asia.* New Holland, London.

Rodriguez Mata, J.R., Erize, F. & Rumboll, M. (2006) *Field guide to birds of South America*. HarperCollins, London.

Rosenberg, G.H. (1990) Habitat specialization and foraging behavior by birds of Amazonian river islands in Northeastern Peru. *The Condor*, **92**, 427–443.
Salisbury, C.L., Seddon, N., Cooney, C.R. & Tobias, J.A. (2012) The latitudinal gradient in dispersal constraints: ecological specialisation drives diversification in tropical birds. *Ecology Letters*, **15**, 847–855.
Schulenberg, T.S., Stotz, D.F., Lane, D.F., O'Neill, J.P. & Parker, T.A. (2010) *Birds of Peru*. Princeton University Press, Princeton.
Sibley, D. (2003a) *Field guide to the birds of Eastern North America*. Christopher Helm, London.
Sibley, D. (2003b) *Field guide to birds of Western North America*. Christopher Helm, London.
Sick, H. (1993) *Birds in Brazil*. Princeton University Press, Princeton.
Simpson, K. & Day, N. (1996) *Field guide to the birds of Australia*. Princeton University Press, Princeton.
Slud, P. (1964) The birds of Costa Rica: distribution and ecology. *Bulletin of the American Museum of Natural History*, **128**, 1–430.
Stiles, F.G. & Skutch, A.F. (1989) *A guide to the birds of Costa Rica*. Christopher Helm, London.
Stotz, D.F., Lanyon, S.M., Schulenberg, T.S., Willard, D.E., Peterson, A.T. & Fitzpatrick, J.W. (1997) An avifaunal survey of two tropical forest localities on the Middle Rio Jiparaná, Rondônia, Brazil. *Ornithological Monographs*, **48**, 763–781.
Svensson, L. (2009) *Collins bird guide, Second edn.* HarperCollins, London.
van Perlo, B. (2006) *Birds: Mexico and Central America.* HarperCollins, London.
van Perlo, B. (2009a) *Field guide to the birds of Southern Africa*. Collins, London.
van Perlo, B. (2009b) *A field guide to the birds of Brazil.* Oxford University Press, Oxford.

Appendix A5.2 Supplementary Results

Table A5.2 The *c* parameter and R^2 values of the power ISAR function for 23 bird habitat island datasets. For each dataset both values are given for the model when fitted to the total compliment of species, and for specialist and generalist species separately. Dataset information is presented in Table S1 in Appendix S1.

Dataset	*C*			R^2		
	All species	Specialists	Generalists	All species	Specialists	Generalists
1	18.53	9.15	7.25	0.28	0.58	0.01
2	5.29	2.46	2.48	0.89	0.88	0.83
3	44.53	5.41	41.59	0.68	0.64	0.41
4	9.55	2.42	7.18	0.95	0.89	0.92
5	52.75	19.23	37.80	0.76	0.78	0.10
6	21.49	1.76	19.70	0.52	0.53	0.45
7	17.66	3.62	13.87	0.90	0.87	0.78
8	6.95	0.35	8.74	0.64	0.88	0.38
9	64.17	13.66	36.38	0.27	0.16	0.21
10	47.02	20.70	26.9	0.54	0.58	0.36
11	21.08	7.95	13.18	0.93	0.90	0.84
12	30.57	5.89	25.81	0.92	0.90	0.93
13	9.99	1.29	9.14	0.77	0.68	0.67
14	6.60	1.12	5.57	0.82	0.79	0.77
15	9.54	0.62	9.11	0.63	0.58	0.48
16	8.75	0.87	8.04	0.74	0.68	0.67
17	16.52	2.09	14.50	0.46	0.32	0.36
18	12.98	1.94	16.18	0.74	0.92	0.08
19	6.54	1.65	4.94	0.81	0.72	0.81
20	7.27	2.82	5.32	0.84	0.86	0.74
21	27.22	8.59	18.65	0.55	0.71	0.40
22	30.09	11.73	18.36	0.41	0.18	0.45
23	9.01	3.71	5.34	0.92	0.82	0.70

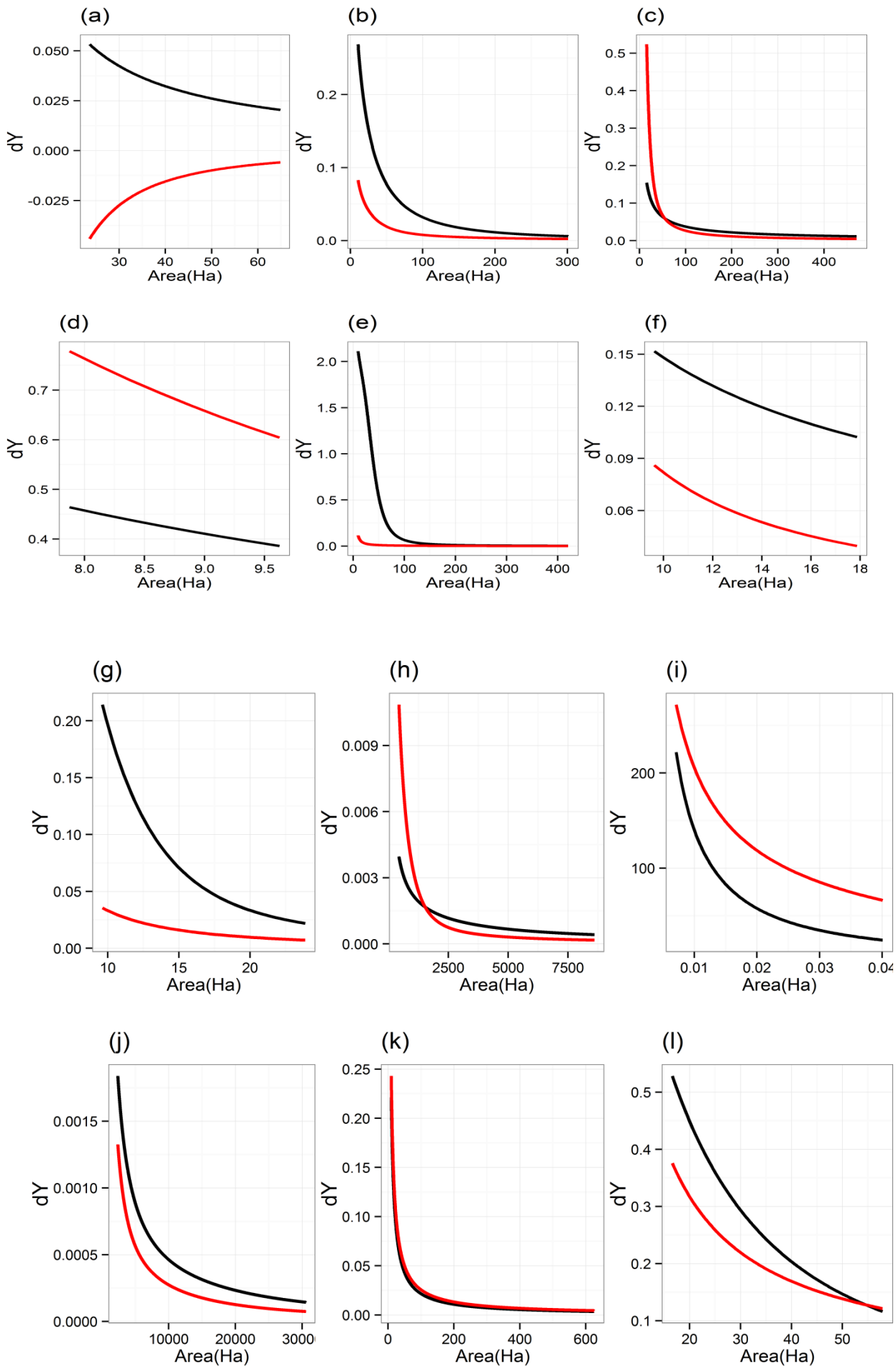
Table A5.3 Model selection results for all 23 bird–habitat island datasets (generalists and specialists presented separately). The data presented are the delta (Δ) AIC_c values. A value of 0 indicates a model provided the best fit accroding to AIC_c and all models within Δ 2 of the best model are considered to be in the best model set (i.e. they are argued to have similar support). Eight models were fit using the 'mmSAR' R package (Guilhamon *et al.,* 2010). 'Expo' refers to the exponential model, and 'Neg Expo' to the negative exponential model. The Weibull model is the three parameter version. NA refers to cases in which the fitting process failed to converge.

Generalists

Dataset	Power	Expo	Neg Expo	Monod	Ratio	Logistic	Weibull	Lomolino
1	0.00	0.28	5.86	5.86	5.11	10.57	10.57	8.18
2	3.36	0.00	13.15	6.80	3.71	4.00	3.83	5.59
3	1.15	0.84	1.14	0.00	5.92	5.81	6.00	6.65
4	4.95	32.36	14.43	5.45	0.86	0.04	0.00	3.96
5	1.09	0.95	8.44	0.00	3.16	4.02	3.62	7.12
6	3.75	2.23	10.49	0.47	2.65	3.20	2.13	0.00
7	17.57	7.57	21.39	0.00	22.51	5.53	10.25	12.75
8	0.18	0.08	0.38	0.00	13.93	13.92	13.90	15.25
9	0.00	0.35	24.99	2.37	4.99	4.34	3.79	3.19
10	0.02	0.00	6.18	1.93	5.72	3.80	3.77	6.02
11	1.23	0.00	11.22	8.86	12.17	9.55	9.64	13.87
12	0.00	1.78	11.80	8.11	28.10	30.00	30.00	28.34
13	4.23	16.49	33.58	25.92	0.00	6.70	6.70	0.81
14	8.96	9.32	13.35	7.16	1.37	4.64	3.50	0.00
15	6.43	15.28	42.28	33.69	0.90	8.88	8.88	0.00
16	1.54	0.00	13.79	6.16	3.46	2.54	2.87	5.83
17	1.19	0.73	1.21	0.00	3.61	3.63	3.53	3.47
18	0.00	0.01	0.18	0.14	NA	NA	NA	NA
19	25.31	10.66	4.73	1.03	2.64	3.43	0.90	0.00
20	0.00	3.16	13.63	11.54	0.39	3.49	3.49	12.39
21	0.00	0.34	7.43	3.39	2.31	7.20	7.20	2.35
22	0.00	0.32	14.02	4.55	4.12	7.20	7.20	3.35
23	0.00	1.74	16.55	10.68	2.08	6.00	6.00	2.09

Specialists

Dataset	Power	Expo	Neg Expo	Monod	Ratio	Logistic	Weibull	Lomolino
1	0.14	0.00	7.70	6.05	1.73	4.48	4.29	10.83
2	7.92	2.36	6.15	0.76	0.00	1.88	1.08	1.96
3	9.79	6.91	0.00	2.87	8.83	8.38	5.99	5.24
4	1.53	26.03	4.43	0.00	2.35	0.47	1.00	5.97
5	8.76	7.79	0.00	1.08	4.04	5.77	3.96	0.70
6	0.48	0.00	2.95	1.39	4.54	3.65	3.65	6.21
7	9.62	15.22	0.00	1.62	7.54	5.64	5.68	6.00
8	0.00	1.89	4.22	2.91	13.68	14.00	14.00	13.89
9	0.98	0.51	13.22	0.00	7.20	4.23	3.82	4.62
10	0.00	0.07	9.64	3.47	2.74	3.82	3.81	6.13
11	5.8	0.00	17.18	12.99	13.09	9.00	8.71	14.93
12	0.00	2.79	7.94	5.47	26.22	30.00	30.00	26.26
13	15.78	25.9	1.83	2.41	4.81	0.90	0.00	4.29
14	1.93	14.6	2.34	0.00	1.26	1.45	1.67	2.24
15	0.33	5.88	0.25	0.00	1.82	2.15	2.12	1.82
16	1.31	6.67	1.14	0.00	1.70	2.11	2.22	2.10
17	3.10	1.81	0.00	0.88	3.99	3.38	3.30	3.35
18	0.94	0.00	3.45	1.12	NA	NA	NA	NA
19	23.12	10.13	0.00	4.05	4.46	2.40	1.35	1.78
20	4.72	10.94	3.29	4.25	2.68	5.30	4.41	0.00
21	0.36	0.00	7.10	2.60	6.40	6.85	6.89	6.23
22	0.00	0.05	17.28	1.74	6.11	7.20	7.20	5.90
23	0.00	0.63	11.14	7.93	6.90	6.00	6.00	7.03


(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)
dY
Area(Ha)

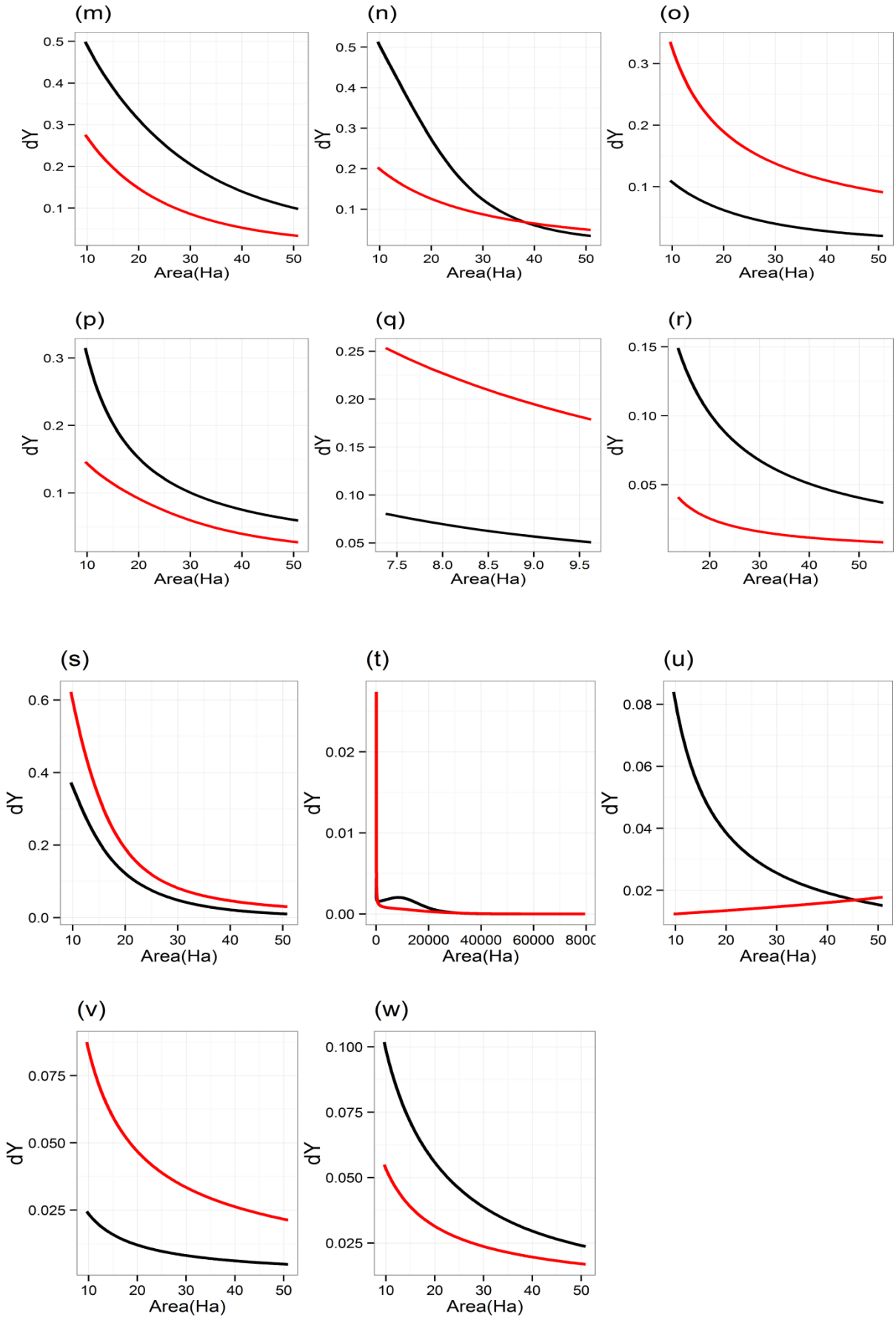
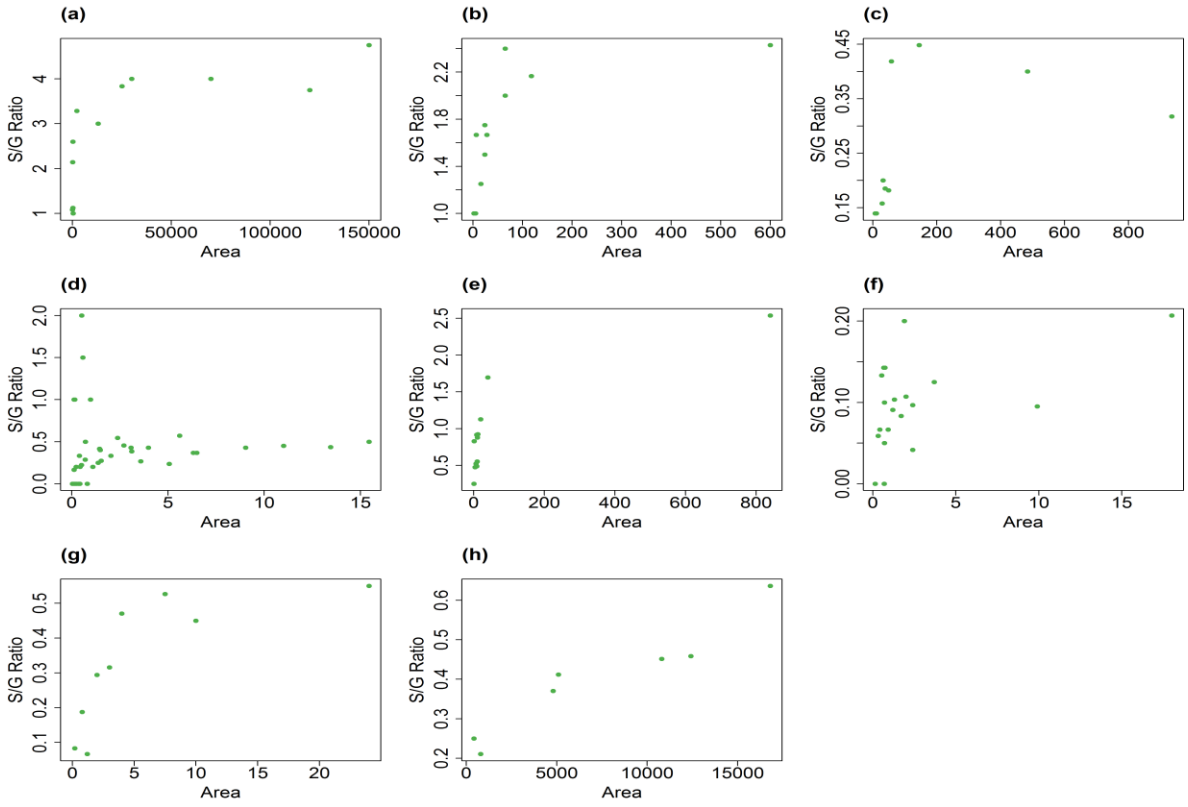

(m)
(n)
(o)
(p)
(q)
(r)
(s)
(t)
(u)
(v)
(w)
dY
Area(Ha)

Figure A5.1 The derivative plots for all 23 bird–habitat island datasets. In each plot the rate of change in species richness predicted by a multi-model SAR, in respect to patch area (dY) is plotted against patch area for habitat specialists (black line) and habitat generalists (red line). Eight SAR models were used to construct the multi-model SAR (power, exponential, negative exponential, Monod, ratio, logistic, Weibull-3) using the 'mmSAR' R package (Guilhaumon *et al.*, 2010). The relative position of the curves indicates the rate at which species richness varies with area. A number of plots have been truncated at the right hand side to enable easier interpretation. The order of the datasets (a-w) follows the order (1-23) in Table A5.1. All areas are in hectares.



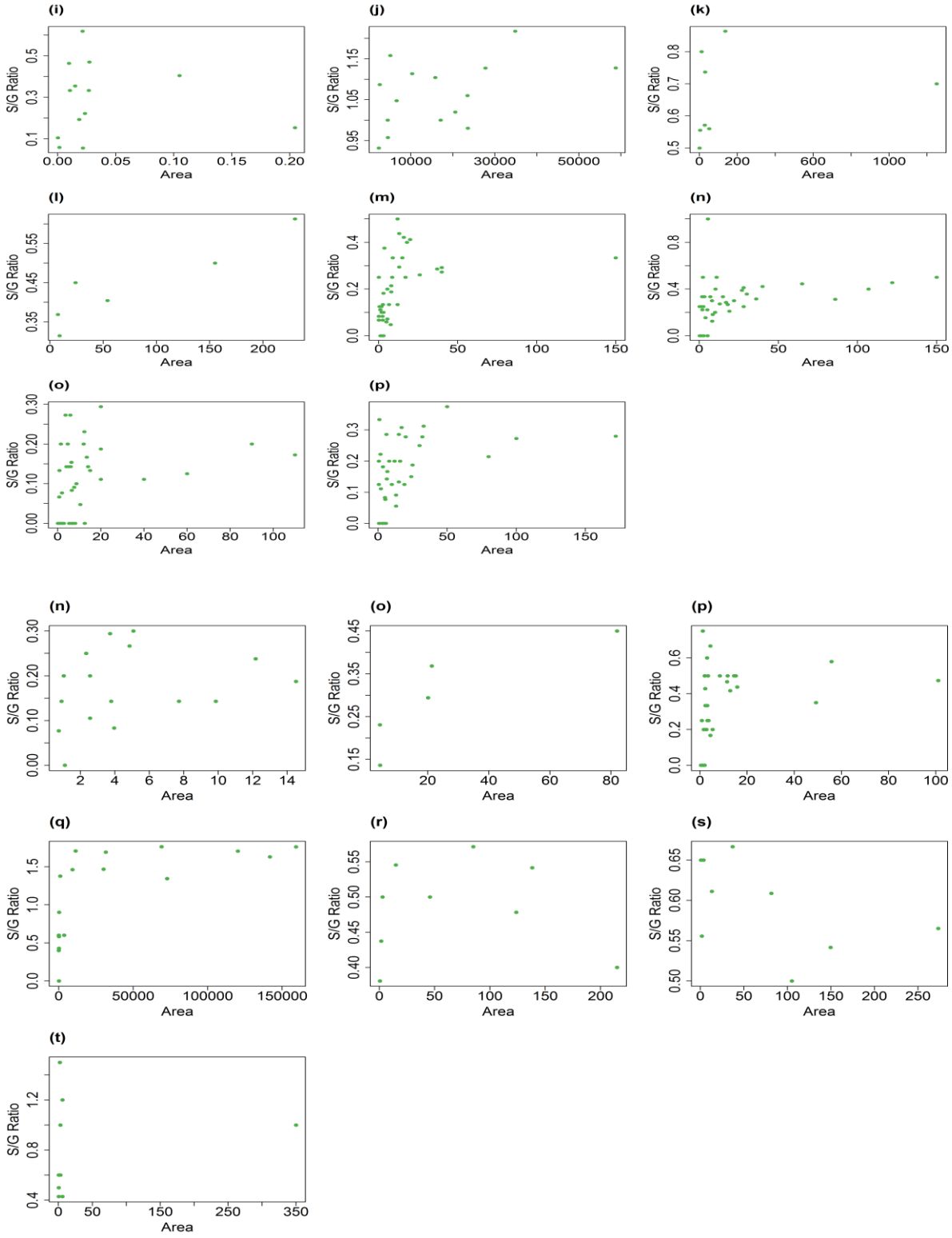


Figure A5.2 The plots of the ratio of specialist species to generalist species (S:G) plotted against habitat island area for all 23 bird habitat island datasets. The order of the datasets (a-w) follows the order (1-23) in Table A5.1.

Appendix 6

Supporting information: Quantifying and interpreting nestedness in habitat islands: a synthetic analysis of multiple datasets (Chapter 8)

Appendix 6.1 Dataset information

Table A6.1 Datasets characteristics for the 97 habitat island datasets used in the analyses, including the taxon, range of island area and range of island richness. The full citations for each dataset are presented after the table. Certain datasets were obtained from the authors of the source papers, whilst others were supplemented with additional data from the source paper authors.

No.	Dataset	Taxon	Area range	Richness range
1	Baldi & Kisbenedek (1999)	Invertebrates	<0.01-40	3-20
2	Behle (1978)	Vertebrates	2849-95312	19-64
3	Benedick et al. (2006)	Invertebrates	120-123027	18-38
4	Blake & Karr (1984)	Vertebrates	2-600	4-24
5	Brotons & Herrando (2001)	Vertebrates	<0.01-38	3-20
6	Brown (1971)	Vertebrates	3108-305102	1-12
7	Brown (1978; birds)	Vertebrates	12173-191142	4-9
8	Brown (1978; mammals)	Vertebrates	3108-305101	3-13
9	Cabrera-Guzmán & Reynoso (2012)	Vertebrates	1-17	7-22
10	Castelletta et al. (2005)	Vertebrates	7-935	49-98
11	Charles & Ang (2010)	Vertebrates	<0.01-149	2-14
12	Cieślak & Dombrowski (1993)	Vertebrates	<0.01-15	1-37
13	Crooks (2002)	Vertebrates	2-102	2-7
14	Crowe (1979)	Plants	<0.01-1	9-70
15	Daily & Ehrlich (1995)	Invertebrates	3-227	5-23
16	Dalecky et al. (2002)	Vertebrates	<0.01-67	1-20
17	Darlington (2001)	Invertebrates	<0.01-0	1-11
18	Davies et al. (2003)	Invertebrates	<0.01-80	39-51
19	Davis et al. (1988)	Vertebrates	17300-2181300	2-10
20	Dickman (1987; scrub)	Vertebrates	<0.01-2	2-13
21	Dickman (1987; woodland)	Vertebrates	<0.01-7	2-17
22	Dinesen et al. (2001) birds	Vertebrates	100-52200	2-8
23	dos Anjos & Boçon (1999)	Vertebrates	<0.01-840	45-138
24	dos Santos et al. (2007)	Plants	12-245	47-110
25	Edenius & Sjöberg (1997)	Vertebrates	2-178	1-17
26	Essl & Dirnböck (2012)	Invertebrates	<0.01-7	5-17

27	Feeley (2003)	Vertebrates	<0.01-180	1-42
28	Fernández-Juricic (2000)	Vertebrates	1-118	6-24
29	Filgueiras et al. (2011)	Invertebrates	10-3500	3-23
30	Flaspohler et al. (2010)	Vertebrates	<0.01-56	2-9
31	Ford (1987)	Vertebrates	<0.01-18	13-35
32	Galle (2008)	Invertebrates	<0.01-0	5-24
33	Galli et al. (1976)	Vertebrates	<0.01-24	3-35
34	Ganzhorn et al. (1999)	Vertebrates	3-600	2-9
35	Gaublomme et al. (2008)	Invertebrates	5-88	20-56
36	Gavish et al. (2012)	Invertebrates	<0.01-4	13-57
37	Gavish et al. (2012)	Invertebrates	<0.01-3	16-62
38	Gillespie & Walter (2001)	Vertebrates	420-16804	21-46
39	Haila et al. (1993)	Vertebrates	1-4	7-19
40	Hatt (1948; amphibians)	Vertebrates	31-15126	1-9
41	Hatt (1948; birds)	Vertebrates	1-15126	2-139
42	Hatt (1948; mammals)	Vertebrates	4-15126	1-13
43	Hattori & Ishida (2000)	Plants	<0.01-8	19-109
44	Holbech (2005)	Vertebrates	2360-58790	72-117
45	Hu et al. (2012)	Plants	<0.01-131	25-121
46	Ishida et al. (1998)	Plants	<0.01-16	18-84
47	Johnson (1975)	Vertebrates	3367-211085	16-60
48	Kelt (2000)	Vertebrates	2-125	3-6
49	Kitchener et al. (1980; lizards)	Vertebrates	34-5119	3-34
50	Kitchener et al. (1980; mammals)	Vertebrates	34-5119	2-13
51	Kratter (1992)	Vertebrates	1-106830	7-48
52	Langrand (1995)	Vertebrates	1-1250	14-51
53	Lomolino & Davis (1997)	Vertebrates	600-2116500	1-15
54	Lomolino & Perault (2001)	Vertebrates	1-59	3-11
55	Lomolino et al. (1989)	Vertebrates	689-1113433	1-16
56	Lumaret et al. (1997)	Plants	<0.01-0	31-102
57	Maldonado-Coelho & Marini (2003)	Vertebrates	4-384	34-52
58	Marini (2001)	Vertebrates	8-230	46-103
59	Matthiae & Stearns (1981)	Vertebrates	<0.01-40	4-13
60	McCollin (1993)	Vertebrates	1-15	12-32
61	Meynard & Quinn (2008)	Vertebrates	180-60830	20-32
62	Matthews et al. (2014; France)	Vertebrates	<0.01-150	5-40
63	Matthews et al. (2014; Spain)	Vertebrates	<0.01-110	6-34
64	Matthews et al. (2014; Norway)	Vertebrates	<0.01-150	2-32

65	Matthews et al. (2014; UK)	Vertebrates	<0.01-172	4-32
66	Miyashita et al. (1998; Tokyo)	Invertebrates	<0.01-27	7-25
67	Miyashita et al. (1998; Yokohoma)	Invertebrates	<0.01-15	15-35
68	Mohd-Azlan & Lawes (2011)	Vertebrates	1-595	30-53
69	Newmark (1991)	Vertebrates	<0.01-521	2-26
70	Nores (1995)	Vertebrates	1300-360000	3-49
71	Nufio et al. (2011)	Invertebrates	1-37	10-29
72	Nyeko (2009)	Invertebrates	10-150	19-36
73	Peltzer et al. (2003)	Vertebrates	<0.01-1	1-18
74	Pineda & Halffter (2004)	Vertebrates	16-72	8-13
75	Ramanamanjato (2000; amphibians)	Vertebrates	10-457	1-20
76	Ramanamanjato (2000; reptiles)	Vertebrates	10-457	5-31
77	Ribas et al. (2005)	Invertebrates	3-299	10-74
78	Rosenblatt et al. (1999)	Vertebrates	2-600	8-14
79	Ruiz-Gutiérrez et al. (2008)	Vertebrates	2-262	62-144
80	Shreeve & Mason (1980)	Invertebrates	2-175	1-22
81	Silva & Porto (2009)	Plants	23-2629	10-54
82	Silva (2001)	Vertebrates	<0.01-6	4-9
83	Simberloff & Martin (1991; forest)	Vertebrates	<0.01-101	2-31
84	Smith et al. (1996)	Vertebrates	<0.01-174	1-17
85	Suarez et al. (1998)	Invertebrates	<0.01-102	1-21
86	Summerville et al. (2002)	Invertebrates	<0.01-0.01	1-18
87	Tonn & Magnuson (1982)	Vertebrates	2-90	1-12
88	Usher & Keiller (1998)	Invertebrates	<0.01-31	68-129
89	Vallan (2000)	Vertebrates	<0.01-1250	9-26
90	Viveiros de Castro & Fernandez (2004)	Vertebrates	1-13	5-11
91	Wang et al. (2010; birds)	Vertebrates	<0.01-1289	15-73
92	Wang et al. (2010; mammals)	Vertebrates	1-1289	2-9
93	Watson (2003)	Vertebrates	2-159246	4-78
94	Weaver & Kellman (1981)	Plants	1-8	8-18
95	Wilson et al. (1994)	Vertebrates	<0.01-350	8-20
96	Yong et al. (2011)	Vertebrates	1-185	11-46
97	Zimmerman & Bierregaard (1986)	Vertebrates	1-500	6-38

Source paper references:

Báldi, A. & Kisbenedek, T. (1999) Orthopterans in small steppe patches: an investigation for the best-fit model of the species-area curve and evidences for their non-random distribution in the patches. *Acta Oecologica*, **20**, 125–132.

Behle, W.H. (1978) Avian biogeography of the Great Basin and intermountain region. *Great Basin Naturalist Memoirs*, **2**, 55–80.

Benedick, S., Hill, J.K., Mustaffa, N., Chey, V.K., Maryati, M., Searle, J.B., Schilthuizen, M. & Hamer, K.C. (2006) Impacts of rain forest fragmentation on butterflies in northern Borneo: species richness, turnover and the value of small fragments. *Journal of Applied Ecology*, **43**, 967–977.

Blake, J.G. & Karr, J.R. (1984) Species composition of bird communities and the conservation benefit of large versus small forests. *Biological Conservation*, **30**, 173–187.

Brotons, L. & Herrando, S. (2001) Reduced bird occurrence in pine forest fragments associated with road proximity in a Mediterranean agricultural area. *Landscape and Urban Planning*, **57**, 77–89.

Brown, J.H. (1971) Mammals on mountaintops: nonequilibrium insular biogeography. *The American Naturalist*, **105**, 467–478.

Brown, J.H. (1978) The theory of island biogeography and the distribution of boreal birds and mammals. *Great Basin Naturalist Memoirs*, **2,** 209–227.

Cabrera-Guzmán, E. & Reynoso, V. (2012) Amphibian and reptile communities of rainforest fragments: minimum patch size to support high richness and abundance. *Biodiversity and Conservation*, **21**, 3243–3265.

Castelletta, M., Thiollay, J.-M. & Sodhi, N.S. (2005) The effects of extreme forest fragmentation on the bird community of Singapore Island. *Biological Conservation*, **121**, 135–155.

Charles, J. & Ang, B. (2010) Non-volant small mammal community responses to fragmentation of kerangas forests in Brunei Darussalam. *Biodiversity and Conservation*, **19**, 543–561.

Cieślak, M. & Dombrowski, A. (1993) The effect of forest size on breeding bird communities. *Acta Ornithologica*, **27**, 97–111.

Crooks, K.R. (2002) Relative sensitivities of mammalian carnivores to habitat fragmentation. *Conservation Biology*, **16**, 488–502.

Crowe, T.M. (1979) Lots of weeds: insular phytogeography of vacant urban lots. *Journal of Biogeography*, **6**, 169–181.

Daily, G.C. & Ehrlich, P.R. (1995) Preservation of biodiversity in small rainforest patches: rapid evaluations using butterfly trapping. *Biodiversity and Conservation*, **4**, 35–55.

Dalecky, A., Chauvet, S., Ringuet, S., Claessens, O., Judas, J., Larue, M.A. & Cosson, J.-F. (2002) Large mammals on small islands: short term effects of forest fragmentation on the large mammal fauna in French Guiana. *Revue d'écologie*, **57**, 145–164.

Darlington, J.P.E.C., Kaib, M. & Brand, R. (2001) Termites (Isoptera) in forest remnants and forest islands in the Shimba Hills National Reserve, coastal province of Kenya. *Sociobiology*, **37**, 527–538.

Davis, R., Dunford, C. & Lomolino, M.V. (1988) Montane mammals of the American Southwest: the possible influence of post-Pleistocene colonization. *Journal of Biogeography*, **15**, 841–848.

Davies, R.G., Hernández, L.M., Eggleton, P., Didham, R.K., Fagan, L.L. & Winchester, N.N. (2003) Environmental and spatial influences upon species composition of a termite assemblage across Neotropical forest islands. *Journal of Tropical Ecology***, 19**, 509–524.

Dickman, C.R. (1987) Habitat fragmentation and vertebrate species richness in an urban environment. *Journal of Applied Ecology*, **24**, 337–351.

Dinesen, L., Lehmberg, T., Rahner, M.C. & Fjeldså, J. (2001) Conservation priorities for the forests of the Udzungwa Mountains, Tanzania, based on primates, duikers and birds. *Biological Conservation*, **99**, 223–236.

dos Anjos, L. & Boçon, R. (1999) bird communities in natural forest patches in southern Brazil. *The Wilson Bulletin*, **111**, 397–414.

dos Santos, K., Kinoshita, L.S. & dos Santos, F.A.M. (2007) Tree species composition and similarity in semideciduous forest fragments of southeastern Brazil. *Biological Conservation*, **135**, 268–277.

Edenius, L. & Sjöberg, K. (1997) Distribution of birds in natural landscape mosaics of old-growth forests in northern Sweden: relations to habitat area and landscape context. *Ecography*, **20**, 425–431.

Essl, F. & Dirnböck, T. (2012) What determines Orthoptera species distribution and richness in temperate semi-natural dry grassland remnants? *Biodiversity and Conservation*, **21**, 2525–2537.

Feeley, K. (2003) Analysis of avian communities in Lake Guri, Venezuela, using multiple assembly rule models. *Oecologia*, **137**, 104–113.

Fernández-Juricic, E. (2000) Bird community composition patterns in urban parks of Madrid: the role of age, size and isolation. *Ecological Research*, **15**, 373–383.

Filgueiras, B.K.C., Iannuzzi, L. & Leal, I.R. (2011) Habitat fragmentation alters the structure of dung beetle communities in the Atlantic Forest. *Biological Conservation*, **144**, 362–369.

Flaspohler, D.J., Giardina, C.P., Asner, G.P., Hart, P., Price, J., Lyons, C.K.a. & Castaneda, X. (2010) Long-term effects of fragmentation and fragment properties on bird species richness in Hawaiian forests. *Biological Conservation*, **143**, 280–288.

Ford, H.A. (1987) Bird communities on habitat islands in England. *Bird Study*, **34**, 205 - 218.

Galle, R. (2008) The effect of a naturally fragmented landscape on the spider assemblages. *North-Western Journal of Zoology*, **4**, 61–71.

Galli, A.E., Leck, C.F. & Forman, R.T.T. (1976) Avian distribution patterns in forest islands of different sizes in central New Jersey. *The Auk*, **93**, 356–364.

Ganzhorn, J.U., Fietz, J., Rakotovao, E., Schwab, D. & Zinner, D. (1999) Lemurs and the regeneration of dry deciduous forest in Madagascar. *Conservation Biology*, **13**, 794–804.

Gaublomme, E., Hendrickx, F., Dhuyvetter, H. & Desender, K. (2008) The effects of forest patch size and matrix type on changes in carabid beetle assemblages in an urbanized landscape. *Biological Conservation*, **141**, 2585–2596.

Gavish, Y., Ziv, Y. & Rosenzweig, M.L. (2012) Decoupling fragmentation from habitat loss for spiders in patchy agricultural landscapes. *Conservation Biology*, **26**, 150–159.

Gillespie, T.W. & Walter, H. (2001) Distribution of bird species richness at a regional scale in tropical dry forest of Central America. *Journal of Biogeography*, **28**, 651–662.

Haila, Y., Hanski, I.K. & Raivio, S. (1993) Turnover of breeding birds in small forest fragments: the "sampling" colonization hypothesis corroborated. *Ecology*, **74,** 714–725.

Hatt, R. (1948) *Island life: a study of land vertebrates of the islands of eastern Lake Michigan.* Cranbrook Institute of Science, Michigan.

Hattori, T. & Ishida, H. (2000) Relationship between species diversity, species composition and forest area of fragmented lucidophyllous forests in central Miyazaki Prefecture. *Japanese Journal of Ecology*, **50**, 221–234.

Holbech, L.H. (2005) The implications of selective logging and forest fragmentation for the conservation of avian diversity in evergreen forests of south-west Ghana. *Bird Conservation International*, **15**, 27–52.

Hu, G., Wu, J., Feeley, K.J., Xu, G. & Yu, M. (2012) The effects of landscape variables on the species-area relationship during late-stage habitat fragmentation. *PloS ONE*, **7,** e43894.

Ishida, H., Hattori, T., Takeda, Y. & Kodate, S. (1998) Relationship between species richness or species composition and area of fragmented lucidophyllous forests in southeastern Hyogo Prefecture. *Japanese Journal of Ecology*, **48,** 1–16.

Johnson, N.K. (1975) Controls of number of bird species on montane islands in the great basin. *Evolution*, **29**, 545–567.

Kelt, D.A. (2000) Small mammal communities in rainforest fragments in central southern Chile. *Biological Conservation*, **92**, 345–358.

Kitchener, D.J., Chapman, A., Dell, J., Muir, B.G. & Palmer, M. (1980a) Lizard assemblage and reserve size and structure in the Western Australian wheatbelt — some implications for conservation. *Biological Conservation*, **17**, 25–62.

Kitchener, D.J., Chapman, A., Muir, B.G. & Palmer, M. (1980b) The conservation value for mammals of reserves in the western Australian wheatbelt. *Biological Conservation*, **18**, 179–207.

Kratter, A.W. (1992) Montane avian biogeography in southern California and Baja California. *Journal of Biogeography*, **19,** 269–283.

Langrand, O. (1995) *The effects of forest fragmentation on bird species in Madagascar: a case study from Ambohitantely Forest Reserve on the central high plateau.* University of Natal, Pietermaritzburg.

Lomolino, M.V. & Davis, R. (1997) Biogeographic scale and biodiversity of mountain forest mammals of western North America. *Global Ecology and Biogeography Letters*, **6,** 57–76.

Lomolino, M.V. & Perault, D.R. (2001) Island biogeography and landscape ecology of mammals inhabiting fragmented, temperate rain forests. *Global Ecology and Biogeography*, **10**, 113–132.

Lomolino, M.V., Brown, J.H. & Davis, R. (1989) Island biogeography of montane forest mammals in the American Southwest. *Ecology*, **70**, 180–194.

Lumaret, R., Guillerm, J.-L., Maillet, J. & Verlaque, R. (1997) Plant species diversity and polyploidy in islands of natural vegetation isolated in extensive cultivated lands. *Biodiversity and Conservation*, **6**, 591–613.

Maldonado-Coelho, M. & Marini, M.Â. (2000) Effects of forest fragment size and successional stage on mixed-species bird flocks in southeastern Brazil. *The Condor*, **102**, 585–594.

Marini, M.Â. (2001) Effects of forest fragmentation on birds of the Cerrado region, Brazil. *Bird Conservation International*, **11,** 13–25.

Matthews, T.J., Cottee-Jones, H.E.W. & Whittaker R.J. (2014) Habitat fragmentation and the species–area relationship: a focus on total species richness obscures the impact of habitat loss on habitat specialists. *Diversity and Distributions*, DOI: 10.1111/ddi.12227

Matthiae, P.E. & Stearns, F. (1981) Mammals in forests islands in southeastern Wisconsin. *Forest Island Dynamics in Man-Dominated Landscapes.* (ed. by R.L. Burgess and D.M. Sharpe), pp. 55–66. Springer-Verlag, New York.

McCollin, D. (1993) Avian distribution patterns in a fragmented wooded landscape (north Humberside, U.K.): the role of between-patch and within-patch structure. *Global Ecology and Biogeography Letters*, **3**, 48–62.

Meynard, C.N. & Quinn, J.F. (2008) Bird metacommunities in temperate South American forest: vegetation structure, area, and climate effects. *Ecology*, **89**, 981–990.

Miyashita, T., Shinkai, A. & Chida, T. (1998) The effects of forest fragmentation on web spider communities in urban areas. *Biological Conservation***, 86,** 357–364.

Mohd-Azlan, J. & Lawes, M.J. (2011) The effect of the surrounding landscape matrix on mangrove bird community assembly in north Australia. *Biological Conservation*, **144**, 2134–2141.

Newmark, W.D. (1991) Tropical forest fragmentation and the local extinction of understory birds in the eastern Usambara Mountains, Tanzania. *Conservation Biology*, **5**, 67–78.

Nores, M. (1995) Insular biogeography of birds on mountain-tops in north western Argentina. *Journal of Biogeography*, **22**, 61–70.
Nufio, C., McClenahan, J. & Deane Bowers, M. (2011) Grasshopper response to reductions in habitat area as mediated by subfamily classification and life history traits. *Journal of Insect Conservation*, **15**, 409–419.
Nyeko, P. (2009) Dung beetle assemblages and seasonality in primary forest and forest fragments on agricultural landscapes in Budongo, Uganda. *Biotropica*, **41**, 476–484.
Peltzer, P.M., Lajmanovich, R.C. & Beltzer, A.H. (2003) The effects of habitat fragmentation on amphibian species richness in the floodplain of the Middle Parana River, Argentina. *Herpetological Journal*, **13**, 95–98.
Pineda, E. & Halffter, G. (2004) Species diversity and habitat fragmentation: frogs in a tropical montane landscape in Mexico. *Biological Conservation*, **117**, 499–508.
Ramanamanjato, J.-B. (2000) Fragmentation effects on reptile and amphibian diversity in the littoral forest of southeastern Madagascar. *Isolated Vertebrate Communities in the Tropics.* (ed. by G. Rheinwald), pp. 297–308. Zoologisches Forschungsinstitut und Museum A. Koenig, Bonn.
Ribas, C.R., Sobrinho, T.G., Schoereder, J.H., Sperber, C.F., Lopes-Andrade, C. & Soares, S.M. (2005) How large is large enough for insects? forest fragmentation effects at three spatial scales. *Acta Oecologica*, **27**, 31–41.
Rosenblatt, D.L., Heske, E.J., Nelson, S.L., Barber, D.M., Miller, M.A. & MacAllister, B. (1999) Forest fragments in east-central Illinois: islands or habitat patches for mammals? *The American Midland Naturalist*, **141**, 115–123.
Ruiz-Gutiérrez, V., Gavin, T.A. & Dhondt, A.A. (2008) Habitat fragmentation lowers survival of a tropical forest bird. *Ecological Applications*, **18**, 838–846.
Shreeve, T.G. & Mason, C.F. (1980) The number of butterfly species in woodlands. *Oecologia*, **45**, 414–418.
Silva, M. (2001) Abundance, diversity, and community structure of small mammals in forest fragments in Prince Edward Island National Park, Canada. *Canadian Journal of Zoology*, **79**, 2063–2071.
Silva, M. & Pôrto, K. (2009) Effect of fragmentation on the community structure of epixylic bryophytes in Atlantic Forest remnants in the northeast of Brazil. *Biodiversity and Conservation*, **18**, 317–337.
Simberloff, D. & Martin, J.L. (1991) Nestedness of insular avifaunas: simple summary statistics masking complex species patterns. *Ornis Fennica*, **68**, 178–192.
Smith, G.T., Arnold, G.W., Sarre, S., Abensperg-Traun, M. & Steven, D.E. (1996) The effect of habitat fragmentation and livestock grazing on animal communities in remnants of gimlet *Eucalyptus salubris* woodland in the western Australian wheatbelt. II. Lizards. *Journal of Applied Ecology*, **33**, 1302–1310.
Suarez, A.V., Bolger, D.T. & Case, T.J. (1998) Effects of fragmentation and invasion on native ant communities in coastal southern California. *Ecology*, **79**, 2041–2056.
Summerville, K.S., Veech, J.A. & Crist, T.O. (2002) Does variation in patch use among butterfly species contribute to nestedness at fine spatial scales? *Oikos*, **97**, 195–204.
Tonn, W.M. & Magnuson, J.J. (1982) Patterns in the species composition and richness of fish assemblages in northern Wisconsin lakes. *Ecology*, **63**, 1149–1166.
Usher, M.B. & Keiller, S.W.J. (1998) The macrolepidoptera of farm woodlands: determinants of diversity and community structure. *Biodiversity and Conservation*, **7**, 725–748.
Vallan, D. (2000) Influence of forest fragmentation on amphibian diversity in the nature reserve of Ambohitantely, highland Madagascar. *Biological Conservation*, **96**, 31–43.

Viveiros de Castro, E.B. & Fernandez, F.A.S. (2004) Determinants of differential extinction vulnerabilities of small mammals in Atlantic forest fragments in Brazil. *Biological Conservation*, **119**, 73–80.
Wang, Y., Bao, Y., Yu, M., Xu, G. & Ding, P. (2010) Nestedness for different reasons: the distributions of birds, lizards and small mammals on islands of an inundated lake. *Diversity and Distributions*, **16**, 862–873.
Watson, D.M. (2003) Long-term consequences of habitat fragmentation—highland birds in Oaxaca, Mexico. *Biological Conservation*, **111**, 283–303.
Weaver, M. & Kellman, M. (1981) The effects of forest fragmentation on woodlot tree biotas in southern Ontario. *Journal of Biogeography*, **8**, 199–210.
Willson, M.F., De Santo, T.L., Sabag, C. & Armesto, J.J. (1994) Avian communities of fragmented south-temperate rainforests in Chile. *Conservation Biology*, **8**, 508–520.
Yong, D.L., Qie, L., Sodhi, N.S., Koh, L.P., Peh, K.S.-H., Lee, T.M., Lim, H.C. & Lim, S.L.-H. (2011) Do insectivorous bird communities decline on land-bridge forest islands in Peninsular Malaysia? *Journal of Tropical Ecology*, **27**, 1–14.
Zimmerman, B.L. & Bierregaard, R.O. (1986) Relevance of the equilibrium theory of island biogeography and species–area relations to conservation with a case from Amazonia. *Journal of Biogeography*, **13**, 133–143.

Appendix 6.2 Random placement model R code and example plots

```
##############################################################################
#############Coleman's Random Placement Model##############################
##############################################################################
##############################################################################
###
#The following code is based on Coleman's (1981) and Coleman et al.'s (1982) random
#placement model. The code is for use with a species-site abundance matrix in which the
#rows are species, and the columns are sites/islands. Each value in the matrix is the
#abundance of a species at a given site. There should be no row or column headings (i.e.
#species or site names). The final row must correspond to the area of each island.
#An example dataset is provided after the code, below.
#The output is a diagnostic plot of island area (log transformed) against island species
#richness. The observed area-richness data points are plotted as solid blue circles. The
#predicted values of the model (solid black line), and one standard deviation (solid red lines)
#are also plotted. To interpret the plot: the model is rejected if more than a third of the
#observed points lie outside the bounds of the standard deviation lines (red lines), and the
#points are not evenly distributed about the standard deviation lines (cf. Wang et al., 2010).
#Implicit within this method is the assumption that species abundances within the set of
#islands under study are an accurate representation of the regional species abundances.
##############################################################################

#First, the two functions which are derived from Coleman et al. (1982), which fit the model
#and determine the standard deviation around the model's predicted values:

sa <- function(x,a){
sa <- (1-x)^a
return(sa)
} # eo sa function
```

```
sa2 <- function(x,a){
sa2 <- (1-x)^(2*a)
return(sa2)
} # eo sa2 function
#The main function, which takes a species-site abundance matrix and produces the diagnostic
#plot:

coleman <-function (data){
##check bottom row cells are all > 0
for (i in 1:ncol(data)){
if(any(data[(nrow(data)),]<=0)){
  stop("Area value <=0")}}

##check each species has 1 or more individuals/each sites as at least one species present
if(any(colSums(data[1:(nrow(data)-1),])==0)){
   warning("Matrix contains sites with no species")}
if(any(rowSums(data[1:(nrow(data)-1),])==0)){
   warning("Matrix contains species which were not sampled")}

####format data######
rowz=nrow(data)
area <- as.vector(unlist(data[rowz,]))
data<- data[1:(rowz-1),]
tot_sp <- nrow(data) #Number of species
tot_area <- sum(area) #Total area

####Derive the observed species richness for each column (site)
ob_sp <- c()
val <- c()
for (j in 1: ncol(data)){
for (i in 1:nrow(data)){
if (data[,j][i] > 0)
          val[i] <- 1
else val[i] <- 0
}#eo i
ob_sp[j] <- sum(val)
}# eo j

###Calculate the relative area of each island
ra <- area/tot_area #relative areas
#get the abundance of each species
abun <- rowSums(data)

##obtain predicted values from random placement model
s_alp <- c()
s_hat <- c()
for (j in 1: length(ra)){
for (i in 1:length(abun)){
s_alp[i] <- sa(ra[j],abun[i])
```

```
} #eo i
s_hat[j] <- tot_sp - sum(s_alp)
} #eo j

##Obtain the model variance
s_alp2 <- c()
s_alp3 <- c()
var <- c()
for (j in 1: length(ra)){
for (i in 1:length(abun)){
s_alp2[i] <- sa(ra[j],abun[i])
s_alp3[i] <- sa2(ra[j],abun[i])
} #eo i
var[j] <-(sum(s_alp2)) - (sum(s_alp3))
} #eo j

##Standard deviation
sd <- sqrt (var)
plus_sd <- s_hat + sd
min_sd <- s_hat - sd

###Plot the model's predictions with standard deviation lines
smoothingSpline = smooth.spline(log(ra), s_hat, spar=0.35)
smoothingSpline2 = smooth.spline(log(ra), plus_sd, spar=0.35)
smoothingSpline3 = smooth.spline(log(ra), min_sd, spar=0.35)
par(adj=0.5)
plot(log(ra),s_hat,xlab="Log(relativearea)",ylab="Species
richness",col="white",cex.lab=1.6,cex=1.6,cex.axis=1.6)
lines(smoothingSpline,col="black",lwd=1.8)
lines(smoothingSpline2,col="red",lwd=1.8)
lines(smoothingSpline3,col="red",lwd=1.8)
points(log(ra),ob_sp,col="blue",pch=19)
}# eo function

#The following is an example species-site matrix (randomly simulated) to use with the above
#code:
```

0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	1	0
1	0	0	0	1	0	0	0	0	0	0
1	0	0	0	0	0	0	0	1	0	0
0	0	0	0	3	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0	1
0	0	3	1	1	4	1	4	3	0	0
0	0	5	0	4	3	5	3	4	5	5
0	0	2	5	2	2	3	0	5	1	3
1	1	0	3	5	1	5	2	4	2	1

0	1	1	1	1	4	1	5	5	4	5
0	1	0	5	4	1	2	1	2	3	5
1	0	0	0	2	0	0	0	0	0	1
0	1	0	0	3	0	0	0	1	1	1
0	2	0	1	3	0	0	0	1	0	1
0	2	1	1	1	0	0	0	1	1	2
0	0	1	0	1	1	1	1	0	5	1
0	0	0	1	3	1	1	1	3	2	0
1	1	1	2	3	0	0	1	2	3	2
2	5	2	3	12	2	4	4	4	8	3
1	3	4	3	10	0	2	1	3	3	2
1.3	2	5.1	8.5	65	1.8	5.5	8.2	10.3	28	4

References

Coleman, B.D. (1981) On random placement and species-area relations. *Mathematical Biosciences*, **54**, 191–215.

Coleman, B.D., Mares, M.A., Willig, M.R. & Hsieh, Y.-H. (1982) Randomness, area, and species richness. *Ecology*, **63**, 1121–1133.

Wang, Y., Bao, Y., Yu, M., Xu, G. & Ding, P. (2010) Nestedness for different reasons: the distributions of birds, lizards and small mammals on islands of an inundated lake. *Diversity and Distribution*s, **16**, 862–873.

Example plots

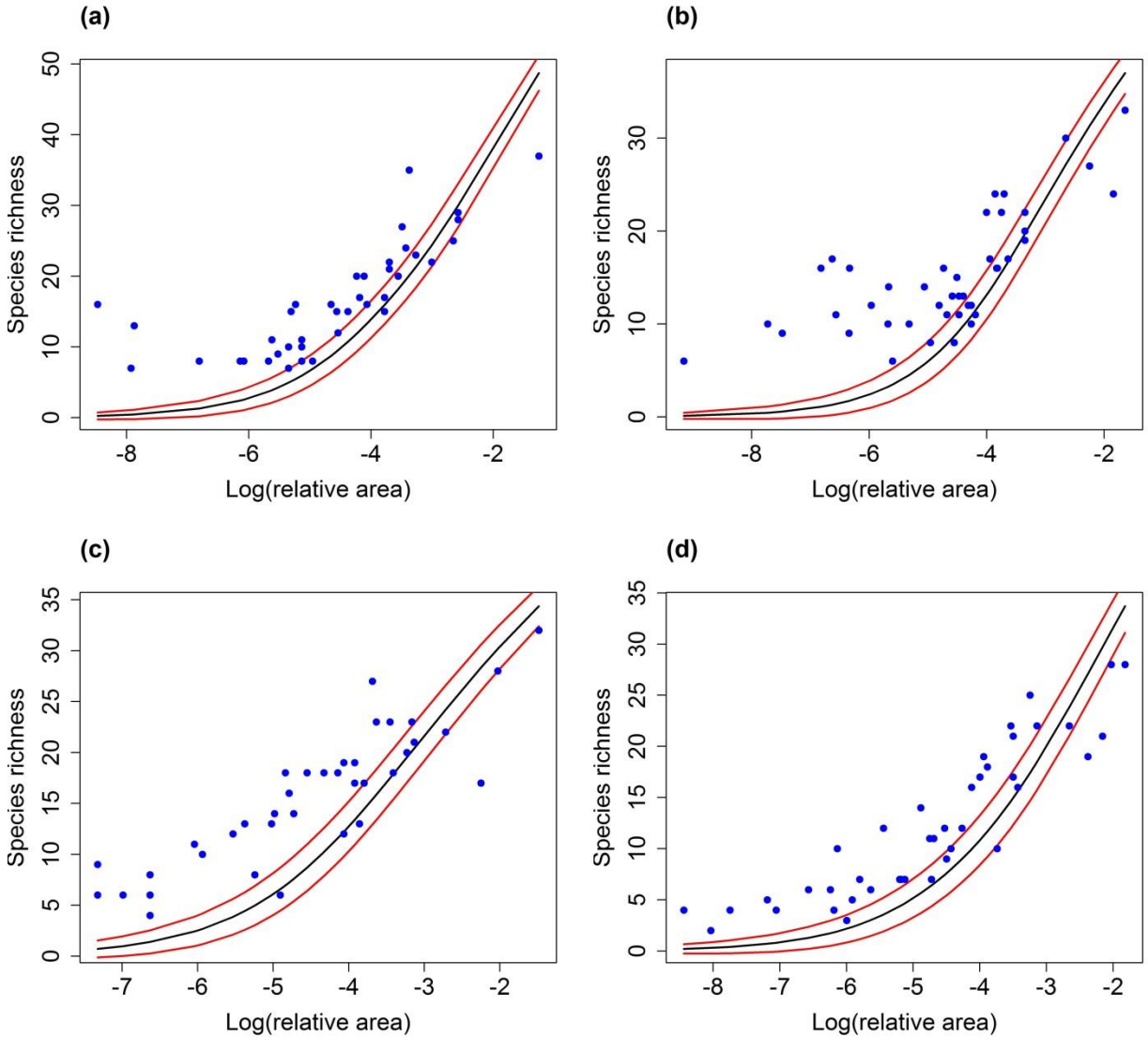


Figure S1 Diagnostic plots of Coleman *et al.*'s (1982) random placement model. In each plot the blue dots represent the data points of the island species–area relationship. The expected data according to the model (black line) and associated confidence intervals (red lines) are also plotted. The random placement model was rejected in each instance if more than a third of the observed data points fell outside the confidence intervals. As such, we rejected the model in all four cases: a) France, b) Spain, c) UK, and d) Norway. The data in each plot (a-d) represents bird species sampled in a set of forest fragments (n= approx. 40 in each case) in an agricultural matrix. Birds were sampled using 10 minute point counts of 50m radius (see Matthews *et al.*, 2014a).

Appendix 6.3 Supplementary results

Table A6.2 The NODF values for the maximally packed matrix, null community simulation results, and the results of our minimum set problem analyses, for 97 habitat island datasets. The significance of the NODF metric values was determined by comparing the observed value with the distribution of values derived from 1000 null communities simulated using the PP model. Significance according to the R00 model was also calculated for comparison. A *Z*-value was calculated as: $Z = (Obs - \mu)/SD$, where Obs is the observed NODF value, μ is the mean nestedness metric value of the null communities (based on 1,000 simulations using the PP algorithm), and SD is the standard deviation of the 1,000 values. A positive *Z*-score and a significant PP null model result indicate that a dataset is significantly nested, whilst a negative *Z*-score and a significant PP null model result indicates that a dataset is significantly anti-nested. *P* values significant at the 0.05 level are highlighted in bold. To determine the solution of the minimum set problem for each dataset we ran an algorithm to calculate the smallest number of habitat islands required in order to include all the species in the dataset. This number was then represented as a proportion of the total number of sites in the dataset (Prop). The proportion of species represented in the largest island (Lar. Prop.) is also given. The dataset numbers correspond to the numbers in Table A6.1

No.	NODF	*Z*-score	PP	R00	Prop. (%)	Lar. Prop. (%)
1	51	0.46	0.32	**<0.01**	22.2	55.6
2	80	0.22	0.41	**<0.01**	35.7	56
3	47	-1.54	**0.03**	**<0.01**	87.5	41.8
4	83	1.77	**0.04**	**<0.01**	16.7	92.3
5	63	-1.34	**0.04**	**<0.01**	19.6	54.1
6	80	1.25	0.11	**<0.01**	11.8	69.2
7	69	-0.6	0.25	**<0.01**	15.4	72.7
8	73	-0.18	0.43	**<0.01**	15.8	62.5
9	62	-0.37	0.35	**<0.01**	83.3	62.9
10	67	-1.15	0.13	**<0.01**	100	62.9
11	50	-1.89	**0.03**	**<0.01**	22.2	77.8
12	62	0.66	0.25	**<0.01**	14.3	74
13	71	0.85	0.2	**0.03**	3.4	28.6
14	48	-0.73	0.23	**<0.01**	57.7	36.7
15	50	-0.46	0.32	**<0.01**	50	52.6
16	57	0.36	0.36	**<0.01**	8.1	71.4

17	40	-0.08	0.49	**<0.01**	26.1	47.8
18	59	-1.65	**0.04**	**<0.01**	70	44.2
19	74	1.01	0.16	**<0.01**	20	83.3
20	72	-0.61	0.25	**<0.01**	33.3	56.2
21	78	0.72	0.24	**<0.01**	25	70
22	82	1.61	0.05	**<0.01**	11.1	66.7
23	66	-0.95	0.17	**<0.01**	91.7	73.4
24	52	-1.88	**0.01**	**<0.01**	92.3	39.1
25	72	2.1	**0.02**	**<0.01**	16.7	60.9
26	69	-1.19	0.12	**<0.01**	11.7	45.5
27	68	1.3	0.1	**<0.01**	7.7	97.7
28	73	-0.28	0.39	**<0.01**	20	46.9
29	56	0.18	0.43	**<0.01**	31.6	76.7
30	85	2.2	**0.01**	**<0.01**	11.1	90
31	70	-2.11	**0.02**	**<0.01**	35	72.9
32	36	-0.45	0.37	**<0.01**	66.7	27.5
33	80	0.03	0.49	**<0.01**	10	100
34	80	0.71	0.24	**<0.01**	28.6	90
35	62	-1.23	0.11	**<0.01**	90	22
36	44	-0.76	0.25	**<0.01**	100	50
37	50	-0.89	0.2	**<0.01**	100	39.1
38	62	-0.89	0.19	**<0.01**	100	52.2
39	62	-1.18	0.11	**<0.01**	46.2	65.5
40	67	0.27	0.4	**<0.01**	22.2	90
41	73	1.88	**0.03**	**<0.01**	33.3	91.4
42	39	-0.73	0.23	**<0.01**	66.7	41.9
43	60	-1.06	0.16	**<0.01**	36.8	61.6
44	76	-1.45	0.07	**<0.01**	53.3	68
45	36	-1.28	0.1	**<0.01**	35.5	20.9
46	64	-0.31	0.38	**<0.01**	31	70.6
47	75	-0.1	0.46	**<0.01**	30	41.9
48	68	-0.27	0.39	**<0.01**	21.4	44.4
49	43	-0.42	0.34	**<0.01**	39.1	31.9
50	60	-0.47	0.33	**<0.01**	30.4	50
51	82	1.43	0.08	**<0.01**	5	100
52	82	2.02	**0.02**	**<0.01**	12.5	100
53	42	-1.4	0.08	**<0.01**	21.7	65.2
54	61	-1.02	0.15	**<0.01**	25	50
55	68	1.07	0.14	**<0.01**	14.8	69.6
56	51	-1.43	0.08	**<0.01**	100	44.6
57	71	-1.94	**0.01**	**<0.01**	66.7	72.2
58	61	-0.62	0.27	**<0.01**	100	71
59	76	0.3	0.38	**<0.01**	4.5	100
60	62	-1.47	0.06	**<0.01**	50	38.8

61	63	-1.93	**0.01**	**<0.01**	66.7	54.9
62	58	-1.95	**0.02**	**<0.01**	27.5	66.7
63	62	-1.73	**0.04**	**<0.01**	19	72.3
64	57	-1.65	**0.04**	**<0.01**	29.3	54.5
65	69	-1.39	0.1	**<0.01**	16.2	74.4
66	67	-0.13	0.46	**<0.01**	57.1	55.9
67	65	-1.17	0.12	**<0.01**	77.8	67.3
68	75	-0.9	0.19	**<0.01**	46.2	54.3
69	66	0.3	0.38	**<0.01**	40	83.9
70	69	0.74	0.23	**<0.01**	33.3	90.7
71	72	0.32	0.37	**<0.01**	38.5	52.6
72	68	-0.87	0.19	**<0.01**	83.3	80
73	76	0.63	0.28	**<0.01**	37.5	90
74	49	-3.54	**<0.01**	0.46	60	61.9
75	88	2.61	**<0.01**	**<0.01**	18.2	95.2
76	75	0.72	0.23	**<0.01**	36.4	75.6
77	48	-0.12	0.45	**<0.01**	61.1	54.4
78	73	-1.79	**0.04**	**<0.01**	30	87.5
79	67	-1.69	**0.03**	**<0.01**	85.7	78.3
80	83	1.19	0.12	**<0.01**	13.6	69.2
81	43	-1.05	0.15	**<0.01**	90	37.4
82	72	-0.62	0.26	**<0.01**	21.4	36.4
83	73	1.91	**0.03**	**<0.01**	14.7	64.4
84	73	1.81	**0.04**	**<0.01**	7.7	89.5
85	53	0.44	0.32	**<0.01**	30	28.9
86	61	0.2	0.41	**<0.01**	32	60
87	47	-0.56	0.28	**<0.01**	33.3	34.8
88	63	-1.76	**0.04**	**<0.01**	66.7	44.4
89	74	-0.19	0.41	**<0.01**	42.9	92.9
90	86	1.48	0.06	**<0.01**	12.5	72.7
91	67	-0.39	0.34	**<0.01**	23.8	78.5
92	66	-0.41	0.33	**<0.01**	21.4	81.8
93	54	-2.9	**<0.01**	**<0.01**	70.6	53.5
94	65	-0.72	0.24	**<0.01**	80	40
95	68	-0.29	0.38	**<0.01**	30	80
96	57	-0.52	0.3	**<0.01**	66.7	61.3
97	80	1.74	**0.04**	**<0.01**	28.6	95

Table A6.3 Values for the nestedness metric based on overlap and decreasing fill (NODF), for 97 habitat island datasets. The metric was calculated for the whole matrix (max), and then separately for matrix rows (Rows; nestedness amongst sites) and matrix columns (Cols; nestedness amongst species incidences). Following Morrison (2013), we took the larger of the row and column values (highlighted in bold) to indicate that a particular type of nestedness contributed more to the overall nestedness pattern.

No.	Max	Cols	Rows	No.	Max	Cols	Rows
1	51	48.3	**55.9**	50	59.8	51.7	**68.6**
2	80.5	80.3	**86.2**	51	81.5	**82.9**	73.6
3	47.4	47.3	**58.6**	52	82.4	82.3	**87.8**
4	82.6	80.8	**91.5**	53	41.5	**46.2**	36.9
5	62.9	52.0	**68.5**	54	61.2	57.1	**64.5**
6	80.3	**82.9**	78.7	55	67.7	58.1	**74.6**
7	69.4	**71.9**	67.7	56	50.8	50.7	**57.2**
8	72.8	70.9	**74.2**	57	70.9	70.9	**73.9**
9	61.7	61.3	**77.7**	58	61.3	61.2	**70.8**
10	67.3	67.2	**73.1**	59	75.9	**81.4**	74.0
11	49.8	44.9	**70.7**	60	62.2	61.5	**69.4**
12	61.8	**62.3**	61.3	61	62.6	62.5	**65.6**
13	70.6	**83.3**	69.9	62	58.1	53.0	**69.6**
14	48.3	47.8	**60.1**	63	62.4	57.5	**68.4**
15	50.1	49.7	**61.8**	64	57.1	51.3	**67.6**
16	56.9	**57.1**	56.8	65	69.1	65.4	**74**.0
17	40.3	38.0	**42.6**	66	66.6	66.3	**75.6**
18	59.0	59.0	**62.5**	67	65.2	65.0	**71.6**
19	73.7	72.7	**75**.0	68	75.1	74.8	**82.8**
20	72.0	70.1	**78.4**	69	66.4	65.5	**76.7**
21	78.2	76.4	**83.4**	70	69.0	68.9	**72.4**
22	81.8	**83.8**	81.3	71	72.2	71.1	**81.7**
23	65.8	65.7	**74.6**	72	67.7	67.5	**85.3**
24	51.6	51.6	**58.3**	73	75.9	**75.9**	75.7
25	72.2	64.9	**84.2**	74	49.1	48.6	**59.9**
26	68.9	56.7	**72.6**	75	87.8	87.3	**89.7**
27	68.4	**71.8**	59.1	76	74.8	73.9	**87.7**
28	73.1	69.3	**79.3**	77	48.3	48.2	**55.2**
29	55.6	54.8	**57.5**	78	72.9	70.6	**78.8**
30	85.3	**86.1**	85.1	79	66.5	66.5	**76.8**
31	70.1	69.1	**76**.0	80	83.4	81.6	**85.8**
32	36.2	35.6	**42.8**	81	43.4	43.2	**56.7**
33	79.6	79.2	**85.3**	82	72.1	**78.2**	68.3
34	80.0	76.7	**87.1**	83	73.2	69.0	**80.6**
35	61.7	61.6	**74.1**	84	73.0	68.7	**75.3**

36	44.4	44.3	**58.6**	85	53.2	38.3	**72.2**
37	50.4	50.3	**65**.0	86	60.7	53.6	**70.9**
38	61.8	61.7	**73**.0	87	46.8	42.6	**53.7**
39	61.7	61.4	**63.3**	88	63.1	63.0	**73.5**
40	67.1	60.9	**74.8**	89	74.4	74.1	**78.7**
41	72.5	72.5	**73.8**	90	86.0	86.0	**86.1**
42	39.1	37.0	**54**.0	91	67.1	64.9	**78.5**
43	60.4	59.9	**73**.0	92	65.9	64.8	**66.5**
44	75.6	75.5	**86.9**	93	53.8	53.8	**55.1**
45	35.7	31.8	**61.1**	94	64.7	63.5	**76.8**
46	63.6	62.8	**77**.0	95	68.4	**68.4**	68.2
47	74.5	74.1	**80.4**	96	56.9	56.9	**69.3**
48	68.3	67.5	**68.6**	97	80.1	79.7	**95.8**
49	43.1	41.4	**58.8**				

Table A6.4 The correlation of the row orders of 97 habitat island presence/absence matrices ordered according to decreasing island area, with the row orders of the maximally packed NODF matrices. Correlation (Corr. Coef.) was determined using Spearman's rank correlation test. Significant *P* values are highlighted in bold.

No.	Corr. Coef.	*P* value	No.	Corr. Coef.	*P* value	No.	Corr. Coef.	*P* value
1	0.72	**<0.01**	36	0.80	**<0.01**	71	0.87	**<0.01**
2	-0.34	0.23	37	0.77	**0.01**	72	0.71	0.14
3	0.67	0.08	38	0.86	**0.02**	73	1.00	**<0.01**
4	0.99	**<0.01**	39	0.69	**0.01**	74	0.9	0.08
5	0.07	0.62	40	0.52	0.16	75	0.99	**<0.01**
6	0.06	0.82	41	0.38	0.16	76	0.93	**<0.01**
7	-0.16	0.60	42	0.97	**<0.01**	77	0.72	**<0.01**
8	0.38	0.11	43	0.96	**<0.01**	78	0.95	**<0.01**
9	1.00	**<0.01**	44	0.63	**0.01**	79	0.82	**0.03**
10	0.83	0.06	45	-0.11	0.17	80	0.75	**<0.01**
11	0.20	0.61	46	0.93	**<0.01**	81	0.47	0.18
12	0.96	**<0.01**	47	0.17	0.47	82	0.39	0.17
13	0.46	**0.01**	48	0.28	0.33	83	0.91	**<0.01**
14	0.34	0.09	49	0.05	0.81	84	0.40	**0.04**
15	0.64	0.10	50	0.26	0.23	85	0.81	**<0.01**
16	-0.02	0.89	51	0.30	0.20	86	0.92	**<0.01**
17	-0.22	0.32	52	0.98	**<0.01**	87	0.62	**0.01**
18	-0.07	0.86	53	0.79	**<0.01**	88	0.40	0.10
19	0.77	**0.01**	54	-0.10	0.66	89	0.96	**<0.01**
20	0.48	0.19	55	0.43	**0.03**	90	0.48	0.24
21	-0.08	0.82	56	-0.14	0.75	91	0.76	**<0.01**
22	-0.05	0.83	57	0.88	**<0.01**	92	0.66	**0.01**
23	0.95	**<0.01**	58	0.94	**0.02**	93	0.85	**<0.01**
24	0.02	0.95	59	0.21	0.35	94	0.52	0.13
25	0.87	**<0.01**	60	0.16	0.55	95	0.88	**<0.01**
26	0.22	0.09	61	0.15	0.71	96	0.89	**0.03**
27	0.70	**<0.01**	62	0.88	**<0.01**	97	0.89	**0.01**
28	0.71	**<0.01**	63	0.74	**<0.01**			
29	0.18	0.45	64	0.19	0.24			
30	0.95	**<0.01**	65	0.16	0.36			
31	0.83	**<0.01**	66	0.86	**0.02**			
32	0.56	**0.03**	67	0.93	**<0.01**			
33	0.99	**<0.01**	68	0.18	0.57			
34	0.68	0.11	69	0.95	**<0.01**			
35	0.25	0.49	70	0.93	**<0.01**			

Table A6.5 The NODF *Z*-scores (of the maximally packed matrix) calculated separately for habitat generalist and habitat specialist bird species subsets, for 17 forest fragment datasets. *Z*-scores were calculated using 1000 null communities simulated with the PP algorithm. Cases in which the specialists' value was the largest are highlighted in bold.

Dataset	Generalist NODF	Specialist NODF
Blake & Karr (1984)	0.98	**2.12**
Castelletta *et al.*, (2005)	-1.08	**-0.74**
Cieślak & Dombrowski (1993)	0.59	-0.28
dos Anjos & Boçon (1999)	-2.04	**1.75**
Ford (1987)	-2.25	**0.51**
Gillespie & Walter (2001)	-0.65	-1.18
Holbech (2005)	-1.46	**-1.38**
Langrand (1995)	1.82	1.72
Marini (2001)	-0.98	**-0.22**
Matthews *et al.* (2014a: France)	-2.64	**0.68**
Matthews *et al.* (2014a: Norway)	-1.5	**0.23**
Matthews *et al.* (2014a: Spain)	-1.82	**-0.01**
Matthews *et al.* (2014a: UK)	-1.47	**1.14**
McCollin (1993)	-1.03	**-0.01**
Simberloff & Martin (1991; forest)	3.05	0.89
Watson (2003)	-2.68	**-2.64**
Willson *et al.* (1994)	-0.31	-0.86

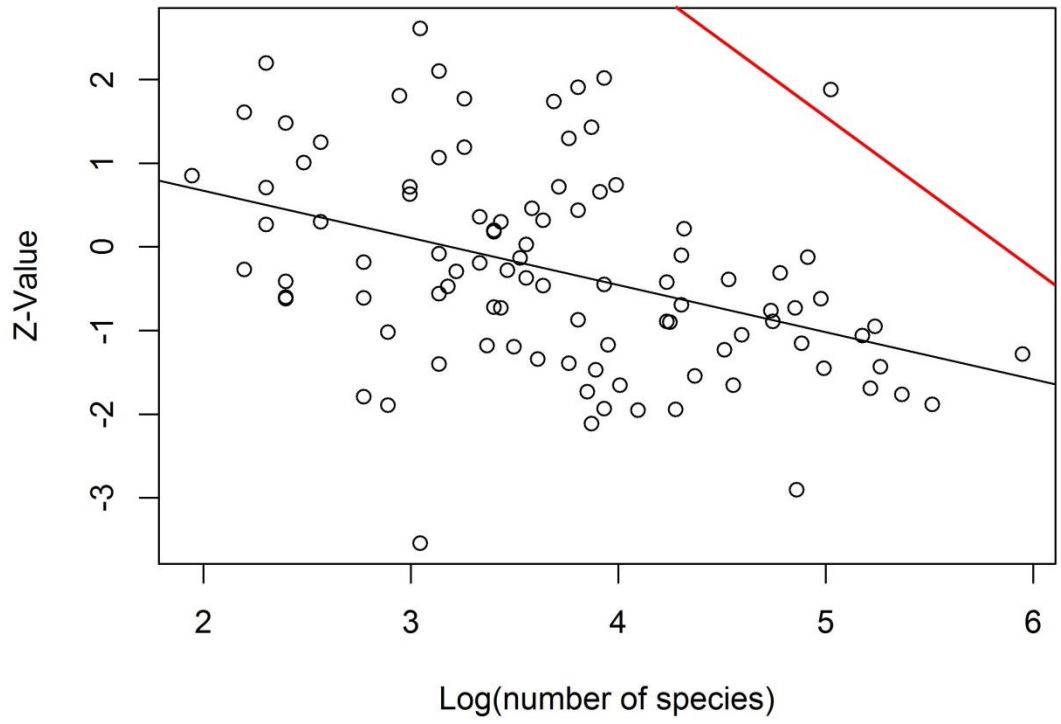


Figure A6.1 The relationship between the number of species in a dataset (log transformed) and the NODF *Z*-score for the maximally packed matrix, for 97 habitat island datasets. The Z-score was calculated by the equation, $Z = (\text{Obs}-\mu)/\text{SD}$, and where Obs is the observed nestedness value according to a given metric, μ is the mean nestedness metric value of the null communities (based on 1,000 simulations), and SD is the standard deviation of the 1,000 values. The PP algorithm was used to simulate the null communities.

Appendix 7

Supporting information: Assessing functional nestedness patterns in habitat islands: an analysis of multiple bird datasets (Chapter 9)

Appendix 7.1 Source paper information

Table A7.1 Description of the source papers, including various dataset characteristics. The number islands has been calculated after removing islands with zero species. + Datasets with species removed due to a lack of trait data.

Study	Location	No. of isl.	No. of sp.	Area range (Ha)	Richness range	Matrix type
1. Blake & Karr(1984)	USA	12	26	1.8–600	4–24	Agriculture
2. Brotons & Herrando (2001)+	Spain	51	37	0.1–38.4	3–20	Agriculture
3. Castelletta *et al.* (2005)+	Singapore	6	131	7–935	49–98	Urban & Grassland
4. Cieślak & Dombrowski (1993)*+	Poland	49	50	0.04–15.44	1–37	Agriculture
5. dos Anjos & Boçon (1999)+	Brazil	12	187	0.5–840	45–137	Grassland
6. Fernández-Juricic (2000)+	Spain	25	31	1–118.2	5–24	Urban
7. Ford (1987)+	UK	20	48	0.14–18	13–35	Agriculture
8. Galli *et al.* (1976)+	USA	10	35	0.01-24	3-35	Agriculture
9. Gillespie & Walter (2001)	Costa Rica	7	69	420–16804	20–45	Agriculture
10. Marini (2001)+	Brazil	6	141	7.5–230	46–103	Agriculture
11. Matthews *et al.* (2014)	France	40	60	0.11–150	5–40	Agriculture
12.Matthews *et al.* (2014)	Norway	41	55	0.2–150	2–32	Agriculture
13.Matthews *et al.* (2014)	Spain	42	47	0.06–110	6–34	Agriculture
14.Matthews *et al.* (2014)	UK	37	43	0.5–172	4–32	Agriculture
15. McCollin (1993)+	UK	16	48	0.74–14.51	12–32	Agriculture
16. Simberloff & Martin (1991)+	Finland	34	45	0.4-101	2-31	
17. Watson (2003)+	Mexico	17	129	2–159242	4–78	Scrub
18. Willson *et al.* (1994)+	Chile	10	25	0.5–350	8–20	Agriculture

Table A7.2 FF null model results for the area-ordered (18 habitat island datasets) matrices. The significance of the traitNODF value (by rows) was determined using the FF null model (1000 permutations), and a *P* value generated from the resultant index values. The topoNODF and S.Fraction significance results are also given. From these values, a Z-transformed score was also generated. The dataset numbers correspond to those in Table A7.1.

No.	Trait (*P*)	Trait (Z)	Topo (*P*)	Topo (*Z*)	S.Frac. (*P*)	S.Frac. (*Z*)
1	0.07	1.49	0.02	1.98	0.13	1.10
2	0.90	-1.26	0.41	0.20	0.97	-1.92
3	0.43	0.27	0.13	1.14	0.63	-0.06
4	0.90	-1.40	0.60	-0.26	0.84	-0.99
5	0.26	0.67	0.07	1.49	0.42	0.26
6	0.23	0.73	0.16	0.99	0.30	0.52
7	0.63	-0.35	0.56	-0.19	0.64	-0.34
8	0.64	-0.29	0.34	0.33	0.7	-0.49
9	1.00	-2.54	0.98	-2.09	0.99	-2.44
10	0.83	-0.88	0.82	-0.91	0.84	-0.74
11	0.96	-1.76	0.41	0.22	0.99	-2.68
12	0.64	-0.34	<0.01	2.71	1.00	-3.15
13	0.34	0.46	0.11	1.22	0.56	-0.13
14	0.97	-1.98	0.32	0.45	1.00	-2.99
15	0.26	0.67	0.43	0.19	0.23	0.77
16	0.84	-1.03	0.24	0.74	0.96	-1.91
17	0.71	-0.49	<0.01	6.38	1.00	-5.52
18	0.02	2.01	<0.01	2.63	0.07	1.52

Appendix 8

Supporting information: The gambin model provides a superior fit to species abundance distributions with a single free parameter: evidence, implementation and interpretation (Chapter 10)

Appendix A8.1 The maximum likelihood estimation of the gambin α parameter, and examples of use of the *gambin* R package.

The maximum likelihood estimation of the gambin α parameter

In order to fit the gambin model, the gamma distribution (i.e. the fitness frequencies) needs to be linked to the actual population sizes. In this context fitness refers to the ecological success of a species, as measured by the chance of achieving a large average population size. Since the extreme species abundance curves are the logseries and lognormal curves, it is convenient to regard the population sizes on a log-scale, i.e. the discrete octaves 0, 1, 2, 3, … where octave 0 relates to 1 individual, octave 1 relates to 2 or 3 individuals, octave 2 relates to 4, 5, 6 or 7 individuals, and so on. By way of an example, let us assume that the most abundant species in the sample is represented by 6719 individuals. Since the 12th octave is the range of individuals from 4096 to 8191, then the most abundant species in the sample is represented by the number 12. Thus, all species in the considered sample are represented by one of the 13 numbers 0, 1, 2, …, 12. In order to take into account abiotic and biotic stochasticity, the assignment of a species' abundance in the 13 available octaves is determined through a binomial process. The problem can then be viewed as linking the gamma distribution with the probability of success in a binomial process with 12 trials.

Since the abundance of any species is always a finite number, it is possible to establish an easy procedure for deducing the probability density for the binomial probabilities by simply cutting the gamma distribution at the 99% percentile. That is, for any value of the shape parameter α (i.e. the community parameter), let $c_{99}(\alpha)$ be the number such that 99% of the density is covered:

$$\int_0^{c99(\alpha)} \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} dx = 0.99$$

We then subdivide the interval $[0, c_{99}(\alpha)]$ into 100 equally long successive pieces, and assume that the fitness of the species takes one of the 100 values $p = p_j = j/100$ (j = 1, 2, …, 100) with probability

$$G_j(\alpha) = \int_{(j-1)c99(\alpha)/100}^{jc99(\alpha)/100} \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} dx / 0.99$$

Note that this function defines the link between the shape parameter α and the probability that the probability of success achieves the value $p = p_j = j/100$ (j = 1, 2, …, 100). If the number of observed species in the samples is denoted S_{OBS}, and the number of species with fitness $p_j = j/100$ is $S_{OBS}*G_j(\alpha)$, then the number of species being represented with an abundance in the k^{th} octave is

$$S_{OBS} * G_j(\alpha) * \binom{12}{k} \left(\frac{j}{100}\right)^k \left(1 - \frac{j}{100}\right)^{12-k} \quad \text{for } k = 0, 1, 2, \ldots, 12.$$

Since species with different fitness frequencies may have population sizes in the same geometric class, the predicted number of species in any octave k is obtained as the sum of the contribution from all possible fitness frequencies:

$\hat{S}_k = S_{OBS} * P_k(\alpha)$, where

$$P_k(\alpha) = \sum_{j=1}^{100} G_j(\alpha) * \binom{12}{k} \left(\frac{j}{100}\right)^k \left(1 - \frac{j}{100}\right)^{12-k} \quad \text{for } k = 0, 1, 2, \ldots, 12$$

Note that $S_{OBS} = \hat{S}_1 + \hat{S}_2 \cdots + \hat{S}_{12}$ where the estimated frequency in each case is determined by only the shape parameter α (since the scale parameter is fixed at 1.0). The unknown parameter may therefore be estimated by maximizing the likelihood function of the sample

$L(\alpha) = P_1^{S_1}(\alpha) * P_2^{S_2}(\alpha) *** P_{12}^{S_{12}}(\alpha)$

The maximum likelihood estimator of the GAMBIN parameter α is the value $\hat{\alpha}_{ML}$ which maximizes the likelihood function L (α).

Examples of the use of the *gambin* package:

data = a vector of abundances of a set of species

Fit the gambin distribution

>fit <- fitGambin(data)

#To obtain α and the associated 95% confidence intervals

> fit$Alpha

> confint(fit)

Plot the distribution

> plot(fit)

Obtain the the chi-square test statistic, degrees of freedom and associated *P* value

> summary(fit)$Chisq

#call the maximum log likelihood value and calculate the various information criterion

> logLik(fit)

>BIC(fit)

>AICc(fit)

#fit the gambin distribution and set 'subsample' to the number of individuals in the least abundant sample (the default value is zero)

>fit <- fitGambin(data, subsample)

Appendix A8.2 Information regarding the Azorean data and sourced datasets.

Table A8.1 Site information for the 18 samples from fragments of native Laurisilva forest from the Azores used in this study. For each fragment the area of the fragment, island in which the fragment is located, and the number of arthropod individuals and number of species sampled are given. Each fragment sample was created by pooling the individuals sampled along four transects in each fragment. Arthropods were sampled using a standardised pitfall trap and canopy beating methodology between 1999 and 2004.

Fragment	Area (Ha)	Island	Number of Individuals	Number of Species
1. Cabeco do Fogo	36	Faial	5586	92
2. Caldeira do Faial	191	Faial	3364	84
3. Lagoas Funda e Rasa	240	Flores	5989	117
4. Morro Alto	1331	Flores	3107	89
5. Caveiro	184	Pico	5578	96
6. Lagoa do Caiado	79	Pico	5766	91
7. Mistério da Prainha	689	Pico	9943	114
8. Pico Pinheiro	74	São Jorge	7366	107
9. Topo	220	São Jorge	4939	102
10. Atalhada	10	São Miguel	3990	130
11. Graminhais	16	São Miguel	2925	84
12. Pico da Vara	306	São Miguel	6204	118
13. Pico Alto	9	Santa Maria	7005	141
14. Biscoito da Ferraria	557	Terceira	5226	116
15. Caldeira Guilherme Moniz	223	Terceira	6027	89
16. Pico Galhardo	38	Terceira	4608	111
17. Serra de Santa Barbara	1347	Terceira	971	68
18. Terra Brava	180	Terceira	4216	103

Table A8.2 Site information for the 6 Azorean island samples used in this study. Island samples were created by pooling the data from the fragments of native Laurisilva forest located on each island (see Table A8.1). For each island the combined area of the pooled fragments, number of pooled fragments, and the number of arthropod individuals and species sampled is given. Arthropods were sampled using a standardised pitfall trap and canopy beating methodology between 1999 and 2004. The island of Santa Maria was not included in the island level analysis as it only possesses a single sampled fragment.

Island	Number of Fragments	Combined Area (Ha)	Number of Species	Total Number of Individuals
1. Faial	2	227	125	8942
2. Flores	2	1571	141	9096
3. Pico	3	952	153	21287
4. São Jorge	2	294	145	12305
5. São Miguel	2	316	178	10194
6. Terceira	4	1788	165	15801

Full acknowledgements for the sourced datasets used in the model comparison.
Under the conditions of use we reproduce the following text (in each case the text was obtained by personal communication with the data holder) associated with those of the datasets used in our analyses that require such statements. For the full references for these datasets, and those which are not included here, see Table 1 in the main paper.

BCI data:
"The BCI forest dynamics research project was made possible by National Science Foundation grants to Stephen P. Hubbell: DEB-0640386, DEB-0425651, DEB-0346488, DEB-0129874, DEB-00753102, DEB-9909347, DEB-9615226, DEB-9615226, DEB-9405933, DEB-9221033, DEB-9100058, DEB-8906869, DEB-8605042, DEB-8206992, DEB-7922197, support from the Center for Tropical Forest Science, the Smithsonian Tropical Research Institute, the John D. and Catherine T. MacArthur Foundation, the Mellon Foundation, the Small World Institute Fund, and numerous private individuals, and through the hard work of over 100 people from 10 countries over the past two decades. The plot project is part of the Center for Tropical Forest Science, a global network of large-scale demographic tree plots."

Hinkley Point Fish data:
P. A. Henderson and Pisces Conservation Ltd kindly provided the data.

Pasoh Forest:
Pasoh 50ha data collection was funded by STRI/CTFS/FRIM and implemented by the Forest Research Institute Malaysia (FRIM) at the Pasoh FRIM Research Station.

Appendix A8.3 Supplementary results

Table A8.3 Simulated lognormal metacommunity results. Four different sets of lognormal metacommunities were simulated, containing a) 50 species, b) 100 species, c) 500 species, and d) 1000 species. Within each of these sets, different metacommunities were simulated, containing a varying number of individuals. There were 9 lognormal metacommunities in total. Five samples were drawn from each metacommunity relating to 1%, 10%, 25%, 50% and 100% of individuals in the metacommunity. For each sample the α parameter of the gambin distribution was calculated, and the maximum octave (maxoct) and modal octave (modal) of the observed abundance distribution (binned into octaves; see methods) were recorded. This process was repeated 100 times and average values were taken.

a)

Fifty Species						
Proportion of community sampled	1000 Individuals			10,000 Individuals		
(%)	α	maxoct	modal	α	maxoct	modal
1	0.03	1	0	2.19	3	0
10	0.43	4	0	6.35	6	2
25	1.43	6	1	9.61	8	3
50	2.26	7	2	13.75	9	4
100	3.00	8	3	18.64	10	6

b)

One Hundred Species						
Proportion of community sampled	1000 Individuals			10,000 Individuals		
(%)	α	maxoct	modal	α	maxoct	modal
1	1.06	1	0	1.14	3	0
10	1.45	3	0	4.15	6	2
25	2.15	4	1	5.97	8	4
50	3.15	5	2	8.84	9	5
100	5.21	6	2	12.69	10	6

c)

Five Hundred Species						
Proportion of community sampled (%)	10,000 Individuals			100,000 Individuals		
	α	maxoct	modal	α	maxoct	modal
1	0.5	2	0	1.15	5	0
10	1.56	4	0	4.88	8	3
25	2.92	5	1	7.03	10	4
50	4.49	6	2	9.75	11	6
100	6.97	7	3	12.79	12	7

d)

Proportion of community sampled (%)	One Thousand Species								
	10,000 Individuals			100,000 Individuals			1,000,000 Individuals		
	α	maxoct	modal	α	maxoct	modal	α	maxoct	modal
1	0.26	1	0	3.38	3	0	2.6	8	2
10	2.29	4	0	3.44	7	2	9.72	11	6
25	2.3	5	1	6.51	8	4	11.79	13	7
50	2.39	6	1	9.6	9	5	14.75	14	8
100	3.17	7	2	13.29	10	6	18.03	15	9

Table A8.4 Simulated logseries metacommunity results. Three different sets of logseries metacommunities (i.e. abundances in the metacommunity were distributed according to a logseries distribution) were simulated, containing a) 50 species, b) 100 species, and c) 500 species. Within each of these sets two different metacommunities were simulated, containing one thousand and ten thousand individuals. Thus, there were six logseries metacommunities in total. Five samples were drawn from each metacommunity relating to 1%, 10%, 25%, 50% and 100% of individuals in the metacommunity. For each sample the α parameter of the gambin distribution was calculated and the maximum octave (maxoct) and modal octave (modal) of the empirical abundance distribution (binned into octaves; see methods) were recorded. This process was repeated 100 times and average values were taken.

a)

	Fifty Species					
Proportion of community sampled	1000 Individuals			10,000 Individuals		
(%)	α	maxoct	modal	α	maxoct	modal
1	0.67	2	0	1.07	5	0
10	0.48	5	0	1.00	8	0
25	0.39	7	0	0.86	9	0
50	0.46	8	0	0.47	10	0
Complete	0.48	9	0	0.48	11	0

b)

	One Hundred Species					
Proportion of community sampled	1000 Individuals			10,000 Individuals		
(%)	α	maxoct	modal	α	maxoct	modal
1	0.18	2	0	1.08	4	0
10	0.51	5	0	0.52	8	0
25	0.39	7	0	0.36	9	0
50	0.39	8	0	0.39	10	0
Complete	0.46	9	0	0.41	11	0

c)

	Five Hundred Species					
Proportion of community sampled	1000 Individuals			10,000 Individuals		
(%)	α	maxoct	modal	α	maxoct	modal
1	0.93	2	0	0.69	4	0
10	0.60	5	0	0.48	7	0
25	0.47	7	0	0.35	9	0
50	0.55	8	0	0.55	9	0
Complete	0.65	9	0	0.64	10	0

Table A8.5 Model selection results for arthropod SADs of six Azorean islands: Faial (1), Flores (2), Pico (3), São Jorge (4), São Miguel (5), and Terceira (6). Arthropods were sampled in native Laurisilva forest using a standardised pitfall trap and canopy beating methodology between 1999 and 2004. For each island the Pearson's χ^2 statistic and associated *P* value (in parentheses) are presented for the gambin, logseries, Poisson lognormal (PLN) and zero-sum multinomial (ZSM) distributions. The Bayesian Information Criterion (BIC) and Akaike's Information Criterion corrected for small sample size (AIC_c) are also given. PLN has two parameters and the ZSM has three parameters. Gambin and the logseries are one parameter models. The best model according to each criterion is highlighted in bold for each island.

	Gambin			Logseries			PLN			ZSM		
	χ^2 (*P*)	BIC	AIC_c	χ^2 (*P*)	BIC	AIC_c	χ^2 (*P*)	BIC	AIC_c	χ^2 (*P*)	BIC	AIC_c
1	**17.0(0.07)**	**553.9**	**553.9**	53.5(0.00)	601.4	601.5	22.7(0.01)	568.6	569.3	20.3(0.01)	586.3	589.4
2	**22.6(0.10)**	**614.2**	**614.3**	46.4(0.00)	648.9	648.9	28.1(0.00)	630.9	631.6	26.9(0.02)	636.8	639.0
3	34.7(0.00)	**724.8**	**724.6**	63.7(0.00)	767.2	767.0	39.4(0.00)	744.4	744.4	**17.7(0.17)**	750.1	751.6
4	**18.4(0.10)**	**628.7**	**628.5**	41.6(0.00)	659.1	658.9	28.7(0.00)	646.2	646.3	87.5(0.00)	661.5	663.8
5	**18.0(0.08)**	**828.5**	**828.5**	40.1(0.00)	868.6	868.5	25.9(0.01)	848.3	848.6	109.5(0.00)	855.6	857.8
6	**32.1(0.01)**	**754.5**	**754.4**	38.0(0.00)	776.9	776.8	34.8(0.00)	775.0	775.0	36.6(0.00)	767.9	769.4

Table A8.6 The proportion of an assemblage/metacommunity required to be sampled in order for the model parameter estimate of the sample to be within specific accuracy thresholds of the parameter estimate of the whole assemblage. The three thresholds are 10% (i.e. the parameter estimate of the sample is within 10% of the assemblage parameter estimate), 25% and 50%. Four metacommunities were simulated, each with 10,000 individuals: 50 species, 100 species, 500 species and 1000 species. For each metacommunity 19 samples were taken, ranging from 5% of individuals in the metacommunity, to 95% of individuals, in intervals of 5%. The three models (gambin, Poisson lognormal, and logseries) were fitted to the samples and the relevant parameters recorded. This process was repeated 100 times and the mean of the parameter values calculated. The proportion of the metacommunity that needed to be sampled in order for the model parameters to be within the three accuracy thresholds was then calculated.

	10% Accuracy	25% Accuracy	50% Accuracy
		50 Species	
Gambin α	55%	50%	15%
PLN mu	60%	30%	5%
LS Alpha	45%	20%	10%
		100 Species	
Gambin α	60%	50%	25%
PLN mu	65%	35%	10%
LS Alpha	60%	30%	15%
		500 Species	
Gambin α	90%	75%	40%
PLN mu	70%	50%	25%
LS Alpha	60%	35%	15%
		1000 Species	
Gambin α	65%	35%	5%
PLN mu	65%	50%	35%
LS Alpha	45%	15%	10%

Table A8.7 Model comparison results for the sample grain size analysis for three fragments of native Laurisilva forest in the Azores (island in parentheses): a) fragment 4 (Flores), b) fragment 7 (Pico), c) fragment 14 (Terceira). For each fragment the model comparison was conducted using the mean abundance of each species across all transects in that fragment, and then for each combination of two transects and so on iteratively up to *n* transects. Comparison was between the gambin distribution, the logseries and the Poisson lognormal distribution (PLN). The best model was selected through a maximum likelihood based comparison comprising gambin, the Poisson lognormal and the logseries distributions. A model was selected as best if it had the lowest BIC and AIC_c values, with the minimum difference in each criterion set at two. Data are for arthropods sampled using a standardised pitfall trap and canopy beating method from 1999 to 2004. The best model for each set of transects is highlighted in bold.

a)

	Gambin		Logseries		PLN	
Transect	BIC	AIC_c	BIC	AIC_c	BIC	AIC_c
1	**219.3**	**219.7**	235.9	236.3	225.5	227.1
2	**268.3**	**268.5**	302.9	303.1	273.5	274.6
3	**306.7**	**306.9**	346.9	347.1	313.5	314.6
4	**313.9**	**314.0**	370.0	370.0	318.8	319.5
5	**379.0**	**379.0**	406.9	407.0	389.8	390.5
6	**384.7**	**384.8**	428.0	428.1	393.9	394.6
7	**392.1**	**392.0**	438.9	438.8	401.5	401.8
8	**397.0**	**396.9**	445.8	445.7	406.9	407.3

b)

	Gambin		Logseries		PLN	
Transect	BIC	AIC_c	BIC	AIC_c	BIC	AIC_c
1	**312.0**	**312.4**	345.4	345.8	320.6	322.2
2	**370.7**	**370.9**	423.8	424.0	379.5	380.6
3	**406.1**	**406.3**	463.3	463.5	415.5	416.6
4	**408.7**	**408.7**	496.3	496.3	416.1	416.8
5	**503.5**	**503.6**	545.3	545.3	516.7	517.4
6	**514.8**	**514.9**	569.7	569.7	526.8	527.5
7	**522.3**	**522.4**	577.8	577.8	533.9	534.6
8	**528.2**	**528.1**	584.1	584.0	541.9	542.3

c)

	Gambin		Logseries		PLN	
Transect	BIC	AIC_c	BIC	AIC_c	BIC	AIC_c
1	**256.2**	**256.8**	283.9	284.5	263.0	265.2
2	**308.1**	**308.5**	354.2	354.5	315.1	316.7
3	**354.5**	**354.7**	404.6	404.8	361.9	363.0
4	**365.1**	**365.3**	427.2	427.4	372.7	373.8
5	**454.7**	**454.9**	478.4	478.6	466.2	467.3
6	**461.5**	**461.6**	504.6	504.7	473.0	473.7
7	**469.5**	**469.6**	516.6	516.7	481.0	481.7
8	**477.6**	**477.6**	523.9	523.9	489.1	489.8

Table A8.8 Simulation results for the pairwise comparison of different binning methods. The bottom left half of each table represents the mean correlation statistic from a Pearson's product moment correlation for each pairwise comparison, and the top right half of each table represents the corresponding standard deviation of the pairwise correlation values. The five bases used to bin the data were loge (natural logarithm), log2, log3, log4 and log5. For each run a community was simulated with a) 10,000 individuals, and b) 100,000 individuals. For the first iteration of the run the number of species was set to 50, and at each subsequent iteration this was increased by 50, up to the maximum of 1000 species (i.e. 20 iterations). At each stage the gambin model was fit to the data using the five different binning methods, and the alpha value recorded. At the end of each run a Pearson's product moment correlation was calculated for each pairwise comparison of alpha sets (i.e. different binning methods) and the values stored. The number of runs was set to 100 and the mean of the correlation values for each pairwise comparison was calculated.

a)

	Loge	Log2	Log3	Log4	Log5
Loge		0.029	0.045	0.056	0.112
Log2	0.909		0.035	0.019	0.109
Log3	0.865	0.927		0.049	0.111
Log4	0.878	0.926	0.803		0.115
Log5	0.880	0.805	0.765	0.807	

b)

	Loge	Log2	Log3	Log4	Log5
Loge		0.068	0.095	0.103	0.165
Log2	0.958		0.099	0.095	0.174
Log3	0.911	0.934		0.090	0.163
Log4	0.905	0.938	0.939		0.164
Log5	0.833	0.842	0.837	0.824	

Supplementary figures

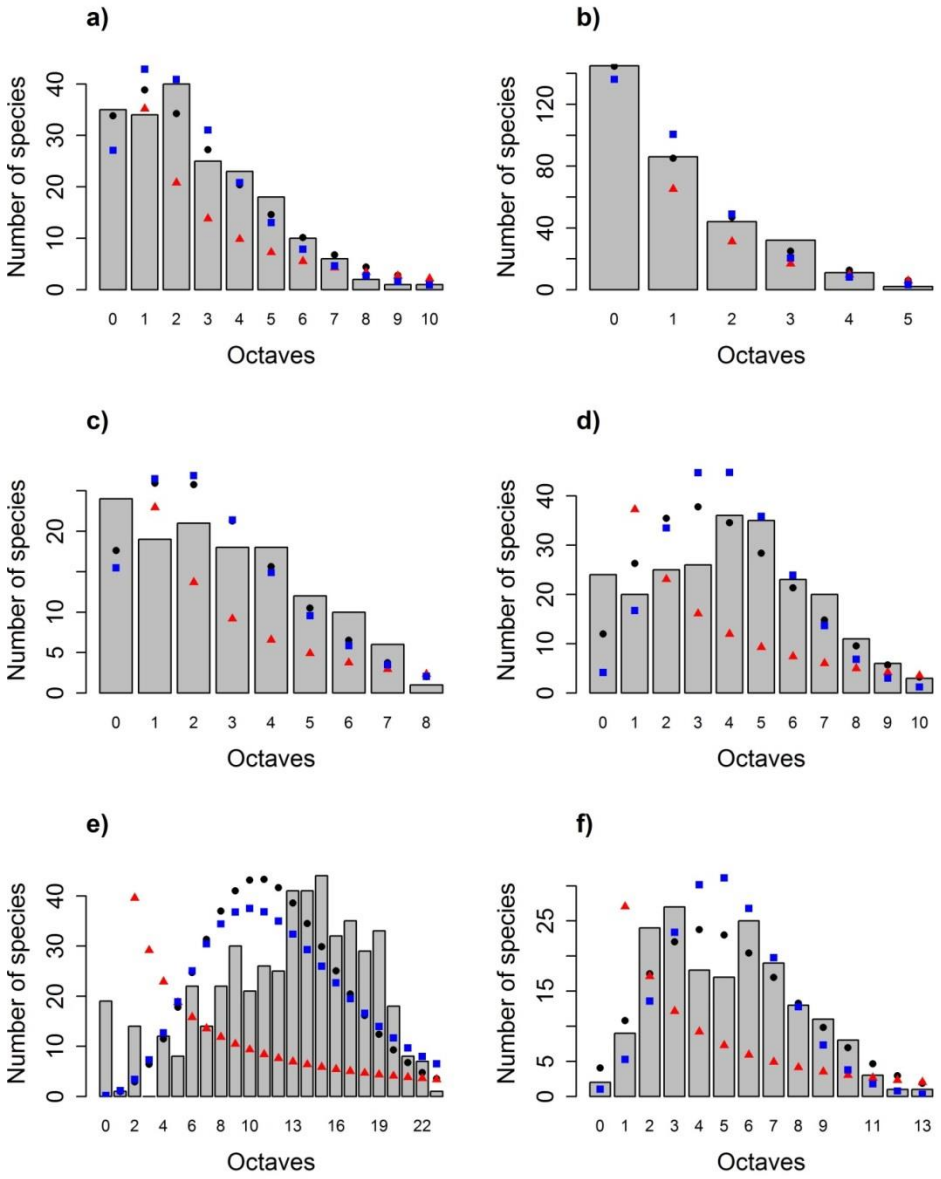


Figure A8.1 Observed (bars) and expected species according to the gambin distribution (black dots), logseries (red triangles) and Poisson lognormal (PLN; blue squares) for a selection of datasets from across the globe. a) Lepidoptera from the Rothamstead monitoring station, UK (α =1.6); b) marine nematodes from the English Channel (α =1.0); c) trees from the Sherman forest plot, Panama (α =2.5); d) trees (>10cm diameter) from the Barro Colorado Island 50ha plot, Panama (α =3.7); e) British breeding birds (α =8.2); f) marine nematodes from the Irish Sea (α =4.3). For reasons of clarity, the predicted values of the logseries have occasionally been omitted when they are much larger than the observed values. For dataset characteristics and references for these datasets, see Table 1 in the main paper.

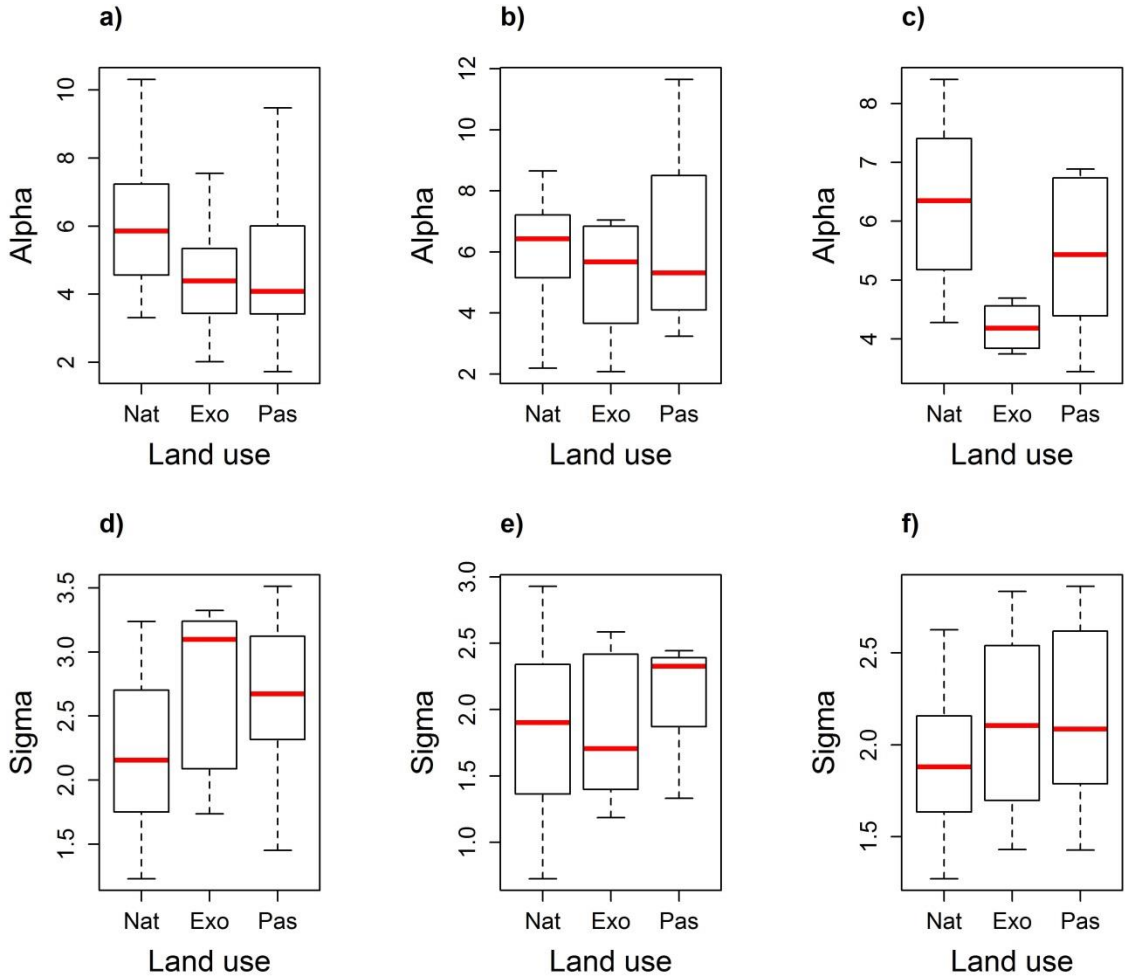


Figure A8.2 Difference in the logseries alpha parameter values (a-c) and the sigma parameter of the Poisson lognormal distribution (d-f) between transects of the three land use types (native forest, exotic plantation forest, and pasture), for three Azorean islands: Terceira (a and d), Faial (b and e), and Flores (c and f). The box plots display the median (red line), the first and third quartiles (black box), and the minimum and maximum values (whiskers). Differences in mean parameter values between land uses on each island were tested for significance using a Wilcoxon rank sum test. The only significant (i.e. $P < 0.05$) difference in alpha values between land uses, were between native forest and pasture on Terceira (a) and native and exotic forest on Flores (c). There were no significant differences in sigma values.

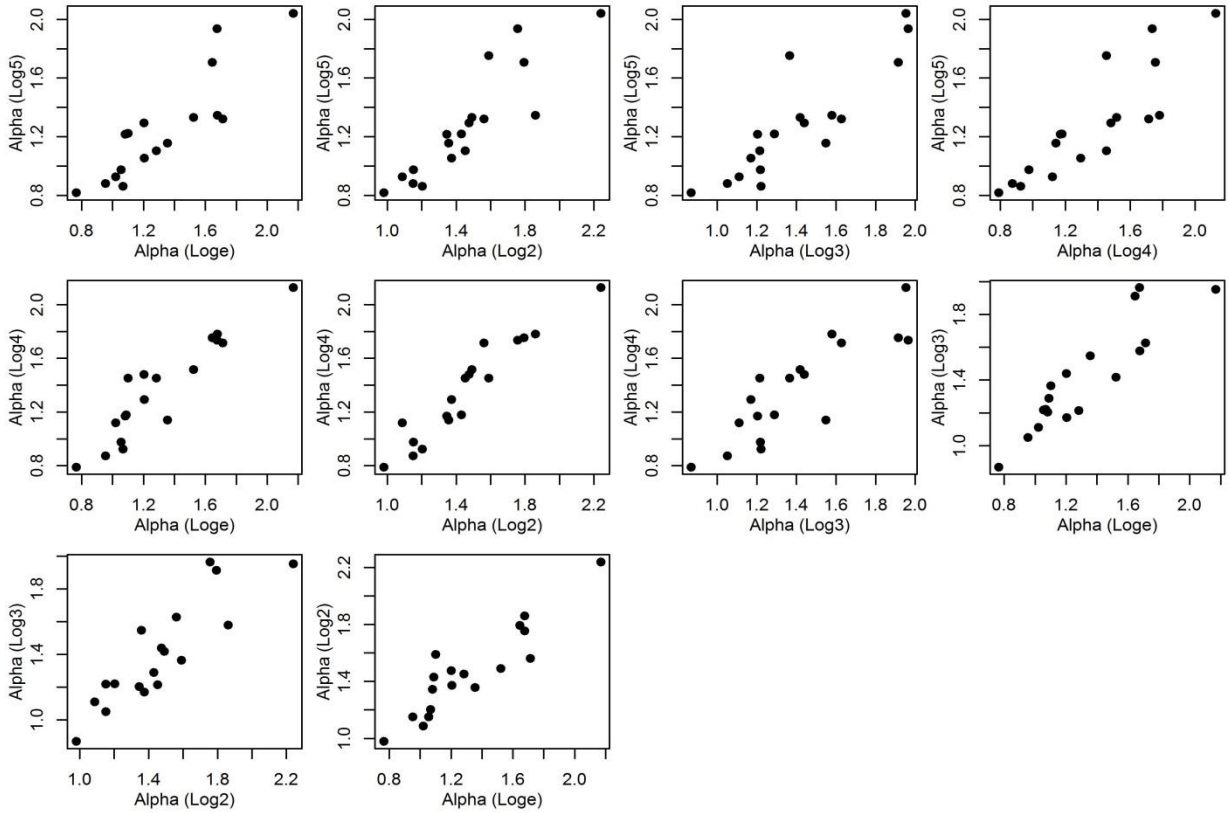


Figure A8.3 The sensitivity of alpha to the binning method employed. Each plot represents a comparison of alpha values derived using a different log base binning method (X and Y axes). The different bases used were loge (natural logarithm), log2 (the main method used), log3, log4 and log5. Each pairwise comparison is significant according to a Pearson's correlation test ($P < 0.001$).

Appendix 9

Supporting information: Multimodal species abundance distributions: a deconstruction approach reveals the processes behind the pattern (Chapter 11)

Table A9.1 Likelihood Ratio Test results for the twelve fragments and five islands of arthropod data from the Azores. LRS= likelihood ratio statistic, P-Value= the P-Value based on a Chi-Squared distribution with the degrees of freedom (df) in parentheses for each set of model comparisons. PLN1, 2, 3= the one, two and three mode Poisson lognormal distributions respectively. As Likelihood Ratio Tests are only applicable when considering nested models the logseries distribution has been excluded.

	FRAGMENTS					
	PLN1/PLN2 (df=3)		PLN2/PLN3 (df=3)		PLN1/PLN3 (df=8)	
Code	LRS	*P*	LRS	*P*	LRS	*P*
1	15.00	0.002	4.97	0.174	19.97	0.003
2	16.91	0.001	2.49	0.477	19.40	0.004
3	6.15	0.104	7.29	0.063	13.44	0.037
4	5.69	0.128	24.47	0.000	30.17	0.000
5	17.52	0.001	1.86	0.603	19.38	0.004
6	20.96	0.000	2.96	0.397	23.93	0.001
7	18.65	0.000	2.53	0.471	21.18	0.002
8	21.15	0.000	42.36	0.000	63.50	0.000
9	16.30	0.001	96.33	0.000	112.63	0.000
10	19.04	0.000	4.35	0.226	23.38	0.001
11	28.00	0.000	5.31	0.151	33.31	0.000
12	22.90	0.000	4.40	0.221	27.30	0.000
			ISLANDS			
Faial	22.90	0.000	2.40	0.494	25.30	0.000
Flores	15.90	0.001	18.90	0.000	34.79	0.000
São Jorge	17.92	0.000	10.67	0.014	28.59	0.000
São Miguel	19.33	0.000	4.08	0.253	23.41	0.001
Terceira	18.11	0.000	3.31	0.347	21.42	0.002

Table A9.2 The model selection results for the core (a) and satellite (b) species subsets at both the fragment and island scale. Data are for arthropod species from the Azores. Both BIC and AIC_c are presented. NA represents a situation in which we were unable to get the model to converge irrespective of the different starting values used in the parameter search. Due to convergence issues the PLN2 was not fit to the satellite subsets at either scale.

a) Core species

FRAGMENTS						
Code	PLN		Logseries		PLN2	
	BIC	AIC_c	BIC	AIC_c	BIC	AIC_c
1	500.56	493.27	493.39	488.48	500.84	486.69
2	494.09	487.14	494.47	489.79	507.26	493.85
3	602.22	594.14	596.33	590.91	613.16	597.35
4	457.32	450.14	452.05	447.21	472.31	458.40
5	645.34	637.90	654.10	649.10	661.02	646.58
6	529.52	522.27	528.84	523.95	533.33	519.27
7	531.65	523.66	525.34	519.98	533.67	518.03
8	611.62	603.83	602.32	597.09	NA	NA
9	549.21	541.58	544.70	539.57	546.75	537.39
10	645.44	637.03	640.64	635.00	654.98	638.46
11	332.56	325.56	319.23	314.52	364.59	351.10
12	607.55	599.45	597.95	592.52	627.82	611.96
ISLANDS						
Faial	706.03	697.72	712.35	706.78	718.33	702.01
Flores	713.34	704.64	722.62	716.79	726.66	709.55
São Jorge	758.22	749.44	770.44	764.55	771.37	754.07
São Miguel	826.49	816.82	833.41	826.93	836.52	817.36
Terceira	879.88	870.61	899.14	892.94	891.71	873.42

b) Satellite species

FRAGMENTS				
Code	PLN		Logseries	
	BIC	AIC_c	BIC	AIC_c
1	214.21	206.91	212.51	207.60
2	125.06	118.11	135.93	131.25
3	279.98	271.91	288.05	282.63
4	216.71	209.53	226.56	221.72
5	158.95	151.52	193.01	188.02
6	207.02	199.76	231.54	226.66
7	398.95	390.96	415.88	410.52
8	304.28	296.49	327.57	322.34
9	126.94	119.31	137.36	132.23
10	175.54	167.13	192.02	186.38
11	67.21	60.22	74.35	69.64
12	177.20	169.10	218.13	212.69
ISLANDS				
Faial	331.37	323.06	337.47	331.90
Flores	415.09	406.39	434.17	428.35
São Jorge	314.47	305.69	362.65	356.76
São Miguel	697.50	687.83	759.44	752.96
Terceira	399.91	390.64	469.79	463.59

Appendix 10

Authorship statements for co-authored papers

In accordance with the regulations set out in the GSO.21 examination notes, letters from each co-author are provided which outline the role the co-authors in the various papers included in this thesis.

SCHOOL OF GEOGRAPHY AND THE ENVIRONMENT
Kostas Triantis
Visiting Research Associate
Oxford University Centre for the Environment, South Parks Road, Oxford OX1 3QY, UK
+44(0)1865 285070
konstantinos.triantis@ouce.ox.ac.uk

Thesis submitted for the degree of Doctor of Philosophy by Thomas J. Matthews

I declare that the manuscripts presented as Chapters 4, 5 & 6 within this thesis, for which I am a co-author, are substantially the work of the first author, Thomas J. Matthews. My role has been limited to help with developing ideas and commenting on drafts. I therefore consent to the inclusion of this material in Tom's doctoral thesis.

I can also confirm that the input of Elli Tzirkalli, whom I co-supervised with Prof. R.J. Whittaker, in Chapter 4 was limited to the provision of datasets.

Prof. Kostas Triantis
Assistant Prof.,
Department of Biology, University of Athens

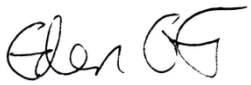


31st October 2014

SCHOOL OF GEOGRAPHY AND THE ENVIRONMENT
Eden Cottee-Jones
DPhil candidate
Oxford University Centre for the Environment,
South Parks Road, Oxford OX1 3QY, UK
+44(0)1865 285070
henry.cottee-jones@ouce.ox.ac.uk; www.geog.ox.ac.uk/research/biodiversity

Thesis submitted for the degree of Doctor of Philosophy by Thomas J. Matthews

I declare that the manuscripts presented as Chapters 7, 8 & 9 within this thesis, for which I am a co-author, are substantially the work of the first author, Thomas J. Matthews. My role has been limited to help with classifying bird species and commenting on drafts. I therefore consent to the inclusion of this material in Tom's doctoral thesis.



Eden Cottee-Jones

31st October 2014

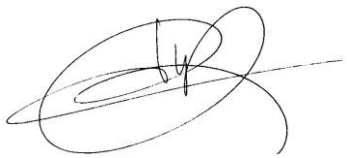
From:
Dr François Guilhaumon
Research scientist
IRD, University of Montpelier

Title

Thesis submitted for the degree of Doctor of Philosophy by Thomas J. Matthews

I declare that the manuscript presented as Chapter 5 within this thesis, for which I am a co-author, is substantially the work of the first author, Thomas J. Matthews. My role has been limited to statistical assistance and commenting on drafts. I therefore consent to the inclusion of this material in Tom's doctoral thesis.

Dr François Guilhaumon



31st October 2014

azorean biodiversity group
http://www.angra.uac.pt/gba

From:
Dr Francois Rigal
Post-Doctoral Researcher
Azorean Biodiversity Group
Universidade dos Açores
Rua Capitão João d`Ávila, São Pedro
9700-042 Angra do Heroísmo, Terceira, Portugal
rigal@uac.pt

Thesis submitted for the degree of Doctor of Philosophy by Thomas J. Matthews

I declare that the manuscripts presented as Chapters 6 & 10 within this thesis, for which I am a co-author, are substantially the work of the first author, Thomas J. Matthews. My role has been limited to statistical assistance and commenting on drafts. I therefore consent to the inclusion of this material in Tom's doctoral thesis.

Dr Francois Rigal



31st October 2014

biogeografie
uni bayreuth

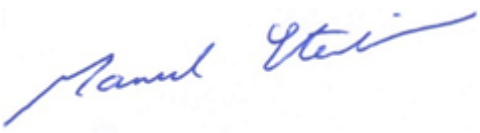
Dr. Manuel Steinbauer Lehrstuhl Biogeografie Universität Bayreuth
95440 Bayreuth

Telefon: +49 (0)921 55---2211 Fax: +49 (0)921 55---23
manuel.steinbauer@uni---bayreuth.de www.uni---bayreuth.de/departments/biogeo
www.global---change---ecology.de

2nd November 2014

Thesis submitted for the degree of Doctor of Philosophy by Thomas J. Matthews

I declare that the manuscripts presented as Chapter 4 within this thesis, for which I am a co--- author, is substantially the work of the first author, Thomas J. Matthews. My role has been limited to statistical assistance and commenting on drafts. I therefore consent to the inclusion of this material in Tom's doctoral thesis.



Dr. Manuel Steinbauer

From:
Prof. Paulo Borges
Group Leader
Azorean Biodiversity Group

azorean biodiversity αβγ group
http://www.angra.uac.pt/gba

Thesis submitted for the degree of Doctor of Philosophy by Thomas J. Matthews

I declare that the manuscripts presented as Chapters 10 & 11 within this thesis, for which I am a co-author, are substantially the work of the first author, Thomas J. Matthews. My role has been limited to providing data and commenting on drafts. I therefore consent to the inclusion of this material in Tom's doctoral thesis.

Prof. Paulo Borges



SCHOOL OF GEOGRAPHY AND THE ENVIRONMENT

Michael K. Borregaard
Postdoctoral Researcher
Oxford University Centre for the Environment,
South Parks Road, Oxford OX1 3QY, UK
Direct tel: +44(0)1865 275845 Reception: +44(0)1865 285070
michael.boreegaard@ouce.ox.ac.uk www.geog.ox.ac.uk/research/biodiversity

Thesis submitted for the degree of Doctor of Philosophy by Thomas J. Matthews

I declare that the manuscripts presented as Chapters 5, 6 & 10 within this thesis, for which I am a co-author, are substantially the work of the first author, Thomas Matthews. My role has been limited to statistical assistance and comments on drafts. I therefore consent to the inclusion of this material in Thomas's doctoral thesis.

Dr Michael Borregaard



31st October 2014

UNIVERSITY OF HELSINKI

LUOMUS
FINNISH MUSEUM OF NATURAL HISTORY

Dr Pedro Cardoso
Curator (Finnish Museum of Natural History)
Associate Professor (University of Helsinki)

Thesis submitted for the degree of Doctor of Philosophy by Thomas J. Matthews

I declare that the manuscripts presented as Chapter * within this thesis, for which I am a co- author, is substantially the work of the first author, Thomas J. Matthews. My role has been limited to statistical assistance and commenting on drafts. I therefore consent to the inclusion of this material in Tom's doctoral thesis.

Dr Pedro Cardoso



31st October 2014

UNIVERSITY OF OXFORD
DEPARTMENT OF ZOOLOGY
Tinbergen Building, South Parks Road, Oxford OX1 3PS

Direct line: (01865) 271 173
e-mail: catherine.sheard@zoo.ox.ac.uk

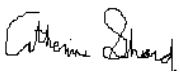
From:
Catherine Sheard
DPhil student
Edward Grey Institute

Title

Thesis submitted for the degree of Doctor of Philosophy by Thomas J. Matthews

I declare that the manuscripts presented as Chapter 9 within this thesis, for which I am a co-author, is substantially the work of the first author, Thomas J. Matthews. My role has been limited to assisting with data collection at the Natural History Museum, Tring, and commenting on drafts. I therefore consent to the inclusion of this material in Tom's doctoral thesis.

Catherine Sheard



4th November 2014

SCHOOL OF GEOGRAPHY AND THE ENVIRONMENT

Robert J Whittaker
Professor of Biogeography
Oxford University Centre for the Environment,
South Parks Road, Oxford OX1 3QY, UK
Direct tel: +44(0)1865 275845 Reception: +44(0)1865 285070
robert.whittaker@ouce.ox.ac.uk www.geog.ox.ac.uk/research/biodiversity

Thesis submitted for the degree of Doctor of Philosophy by Thomas J. Matthews

I declare that the manuscripts presented as Chapters 4-12 within this thesis, for which I am a co-author, are substantially the work of the first author, Thomas Matthews and that author contributions statements provided in submitted and published versions of the chapters accurately describe the contributions of co-authors. While practical assistance, data, and/or statistical guidance was variously provided to Tom by myself, Karl Ugland, and the other co-authors who have each provided similar statements, Tom has in each case taken the lead in the research, analysis and writing.

Professor Robert Whittaker

Robert J Whittaker

07th December 2014

Appendix 11

Dispensation letter for exceeding the allowed thesis word count

ACADEMIC OFFICE

Social Sciences Divisional Office,
Hayes House,
75 George Street,
Oxford
OX1 2BQ

Mr Thomas Matthews
5 Manor Close
Shrivenham
Swindon
SN6 8AE

Ref. GSO

OSS ID: 490700

3 December 2014

Dear Mr Matthews

I am writing to confirm that you have been granted permission, by the School of Geography's, Director of Graduate Studies, to exceed the word limit for your DPhil thesis, by excluding the appendices from the word count.

Yours sincerely

AM Smith

Mrs A M Smith
Graduate Studies Administrator

Cc Professor R. Whittaker, School of Geography, South Parks Road
Tutor for Graduates, Linacre College

General Enquiries Tel: +44 (0)1865 270000 Direct Line Tel: +44 (0)1865 614861

Fax: +44 (0)1865 614855 Email: mary.smith@admin.ox.ac.uk Web: www.socsci.ox.ac.uk