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

Matthews, Thomas; Guilhaumon, Francois; Triantis, Kostas A. ; Borregaard, Michael K.; Whittaker, Robert J.

DOI:
10.1111/geb.12269

Article Type: Meta-analysis

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

Thomas J. Matthews[1,2], François Guilhaumon[2,3], Kostas A. Triantis[1,2,4], Michael K. Borregaard[1] & Robert J. Whittaker[1, 5]

[1]Conservation Biogeography and Macroecology Programme, School of Geography and the Environment, University of Oxford, South Parks Road, Oxford, OX1 3QY, UK

[2]Azorean Biodiversity Group, Departamento de Ciências Agrárias – CITAA, Universidade dos Açores, Angra do Heroísmo, Pico da Urze, 9700-042, Terceira, Açores, Portugal

[3]IRD Laboratoire Écologie des Systèmes Marins Côtiers UMR 5119, CNRS, IRD, IFREMER, UM2, UM1, cc 093, Place E. Bataillon, 34095 Montpellier Cedex 05, France

[4]Department of Ecology and Taxonomy, Faculty of Biology, National and Kapodistrian University, Athens GR-15784, Greece

[5]Center for Macroecology, Evolution and Climate, Natural History Museum of Denmark, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen Ø, Denmark

*Correspondence: Thomas J. Matthews, Conservation Biogeography and Macroecology Programme, School of Geography and the Environment, University of Oxford, South Parks Road, Oxford, OX1 3QY, UK

Email: thomas.matthews@ouce.ox.ac.uk

Running header: Species–area relationships in islands and habitat islands

Word count: abstract (292): 5511 words main text: 4 tables; 3 Figures; 48 references; 3 appendices

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.

Keywords Applied island ecology, conservation biogeography, fragmentation, habitat islands, habitat loss, island biogeography, island species–area relationship, macroecology, multi-model comparison, species–area relationship

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 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.

MATERIALS AND METHODS

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 S1 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.

Model comparison

Following Triantis *et al.* (2012) we compared 20 ISAR models (Table 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).

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.

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 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.

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 S1 in Appendix S2).

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).

RESULTS

From >1500 published articles, 207 habitat island datasets (Appendix S2) 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. S1 in Appendix S2). 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.

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 S1 in Appendix S2). 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. 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. 1, Table 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. 1). The more complex models performed poorly, with the weibull4, betap, and heleg models (Table 1) never providing the best fit (Fig. 1). The linear model ranked second in the number of best fits (Fig. 1a), but was ranked lower in generality and efficiency (Fig. 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 S1).

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. 2). The main exception was the linear model (Fig. 2), the rank of which significantly decreased with increasing minimum *Ni* (Fig. 2).

The model rankings appeared relatively robust to the removal of individual data points (Table S2 in Appendix S3). 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%.

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 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.

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. 3), and the median *c* value was 2.27 (Q1 & Q3 = 1.14 & 3.03; Fig. S2a in Appendix S3). In comparison, *z* values were larger for continental-shelf islands (median=0.28; Q1 & Q3= 0.19 & 0.37; Fig. 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. S3 in Appendix S3). 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. S2a in Appendix S3), 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. 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. S2b in Appendix S3), but not urban islands (P = 0.11). The median *z* values did not significantly differ between taxa (Fig. 3c), while vertebrates had a significantly lower *c* value than plants (Wilcoxon tests; P = <0.01; Fig. S2c in Appendix S3), and invertebrates (P = 0.01).

Considering only significant *z* values, when *z* was used as the response variable in a set of LMs, the best model contained *Ascale* (Fig. 3d), *Ni*, *Smin* and *Smax* (Table 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. S4 in Appendix S3). 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 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 S3 and Figs. S5 & S6 in Appendix S3). For simplicity, in the Discussion, we focus solely on results based on significant parameter estimates.

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.

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. 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 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 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 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).

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 4).

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 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. 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. 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.

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.S7 in Appendix S3 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 (Matthews *et al.*, submitted); (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 S2 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.

ACKNOWLEDGEMENTS

François Rigal and three anonymous reviewers provided extremely useful comments. Alison Pool kindly helped with data entry. A number of authors kindly provided data. TJM acknowledges funding from the Royal Geographical Society, the Sir Richard Stapley Trust, and the Sidney Perry Foundation. RJW acknowledges funding from the Biogeography Research Group of the Royal Geographical Society (with IBG) for attendance at Intecol 2014, where a part of this work was presented. MKB was supported by an individual Sapere Aude postdoctoral grant from the Danish Councils for Independent Research.

REFERENCES

Anderson, M.J. & Willis, T.J. (2003) Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. *Ecology*, **84**, 511-525.

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.

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.

Bartoń, K. (2012) *MuMIn: multi-model inference.* R package version 1.2. Available at: http://cran.r-project.org/ (accessed 15 September 2013).

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.

Brown, J.H. (1971) Mammals on mountaintops: nonequilibrium insular biogeography. *American Naturalist*, **105**, 467-478.

Burnham, K.P. & Anderson, D.R. (2002) *Model selection and multi-model inference: a practical information-theoretic approach*, 2nd edn. Springer, New-York.

Connor, E.F. & McCoy, E.D. (1979) Statistics and biology of the species-area relationship. *American Naturalist*, **113**, 791-833.

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.

Dengler, J. (2009) Which function describes the species–area relationship best? A review and empirical evaluation. *Journal of Biogeography*, **36**, 728-744.

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.

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.

Gould, S.J. (1979) An allometric interpretation of species-area curves: the meaning of the coefficient. *The American Naturalist*, **114**, 335-343.

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.

Guilhaumon, F., Mouillot, D. & Gimenez, O. (2010) mmSAR: an R-package for multimodel species–area relationship inference. *Ecography*, **33**, 420-424.

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.

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.

He, F. & Legendre, P. (1996) On species-area relations. *The American Naturalist*, **148**, 719-737.

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., 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.

Laurance, W.F. (2008) Theory meets reality: how habitat fragmentation research has transcended island biogeographic theory. *Biological Conservation*, **141**, 1731-1744.

Losos, J.B. & Schluter, D. (2000) Analysis of an evolutionary species-area relationship. *Nature*, **408**, 847-850.

MacArthur, R.H. & Wilson, E.O. (1967) *The theory of island biogeography*. Princeton University Press, Princeton.

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., Cottee-Jones, H.E.W. & 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 & Distributions*, DOI: 1320408-1267725.

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.

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.

Pereira, H.M. & Daily, G.C. (2006) Modelling biodiversity dynamics in countryside landscapes. *Ecology*, **87**, 1877-1885.

Preston, F.W. (1962) Canonical distribution of commoness and rarity.1. *Ecology*, **43**, 185-215.

R Development Core Team (2013) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.r-project.org/ (accessed 13 August 2013).

Rosenzweig, M.L. (1995*) Species diversity in space and time*. Cambridge University Press, Cambridge.

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.

Scheiner, S.M. (2003) Six types of species-area curves. *Global Ecology and Biogeography*, **12**, 441-447.

Scheiner, S.M., Chiarucci, A., Fox, G.A., Helmus, M.R., McGlinn, D.J. & Willig, M.R. (2010) 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.

Smith, A.B. (2010) Caution with curves: caveats for using the species-area relationship in conservation. *Biological Conservation*, **143**, 555-564.

Sólymos, P. & Lele, S.R. (2012) Global pattern and local variation in species–area relationships. *Global Ecology and Biogeography*, **21**, 109-120.

Tjørve, E. (2009) Shapes and functions of species–area curves (II): a review of new models and parameterizations. *Journal of Biogeography*, **36**, 1435-1445.

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., 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., and 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.

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.

Watling, J.I. & Donnelly, M.A. (2006) Fragments as islands: a synthesis of faunal responses to habitat patchiness. *Conservation Biology*, **20**, 1016-1025.

Whittaker, R.J. (2010) Meta-analyses and Mega-mistakes: calling time on meta-analyses 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. & Matthews, T.J. (2014) The varied form of species–area relationships. *Journal of Biogeography*, **41**, 209-210.

Williams, M.R., Lamont, B.B. & Henstridge, J.D. (2009) Species–area functions revisited. *Journal of Biogeography*, **36**, 1994-2004.

Additional references to the data sources used in this study can be found in Appendix S2 at [URL].

SUPPORTING INFORMATION

Appendix S1. Expanded study methodology.

Appendix S2. Dataset source information.

Appendix S3. Supplementary results

BIOSKETCHES

Tom Matthews is a doctoral student at the University of Oxford. He is interested in the study of macroecological patterns in habitat islands, and the application of island theory to fragmented landscapes. His recent work has focused on species abundance distributions in fragmented landscapes, and the use of piecewise regression to detect thresholds in the species–area relationship.

Author contributions: 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.

TABLES

Table 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

Table 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	189	53.93		

Table 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		

Table 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.

FIGURES

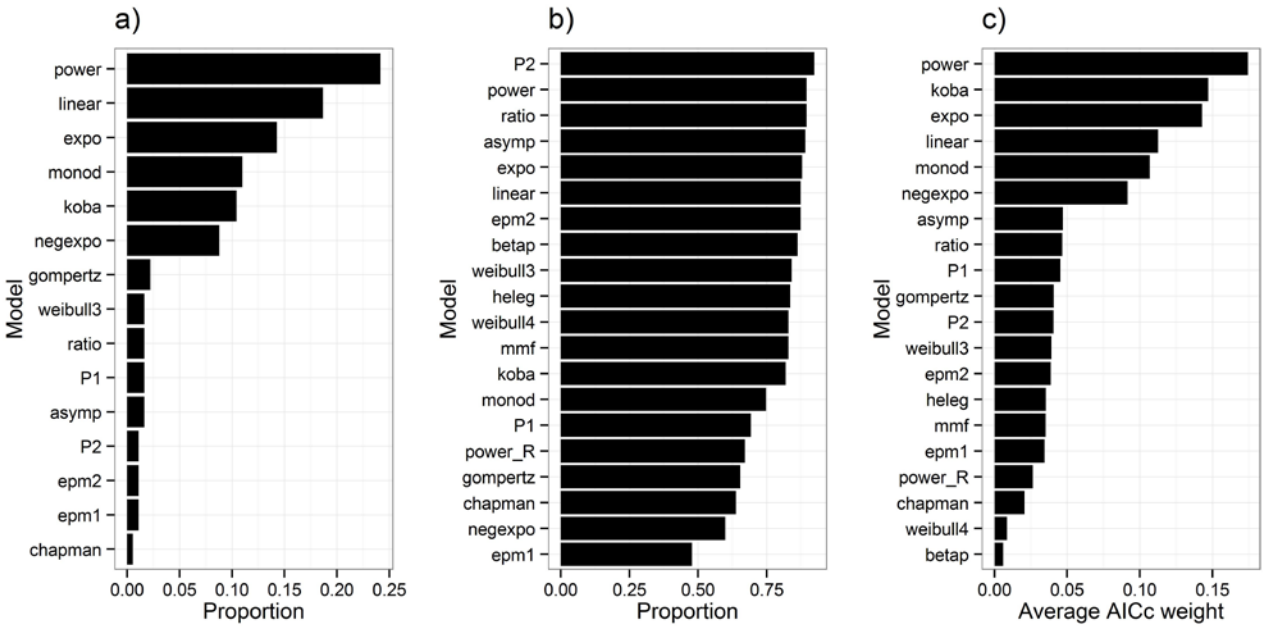

Figure 1 The performance of 20 species–area relationship models fitted to 182 habitat island datasets. Performance was measured in three ways: (a) the proportion of datasets for which a given model provided the best fit (i.e. had the lowest AIC_c value), (b) the proportion of datasets for which a given 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 1.

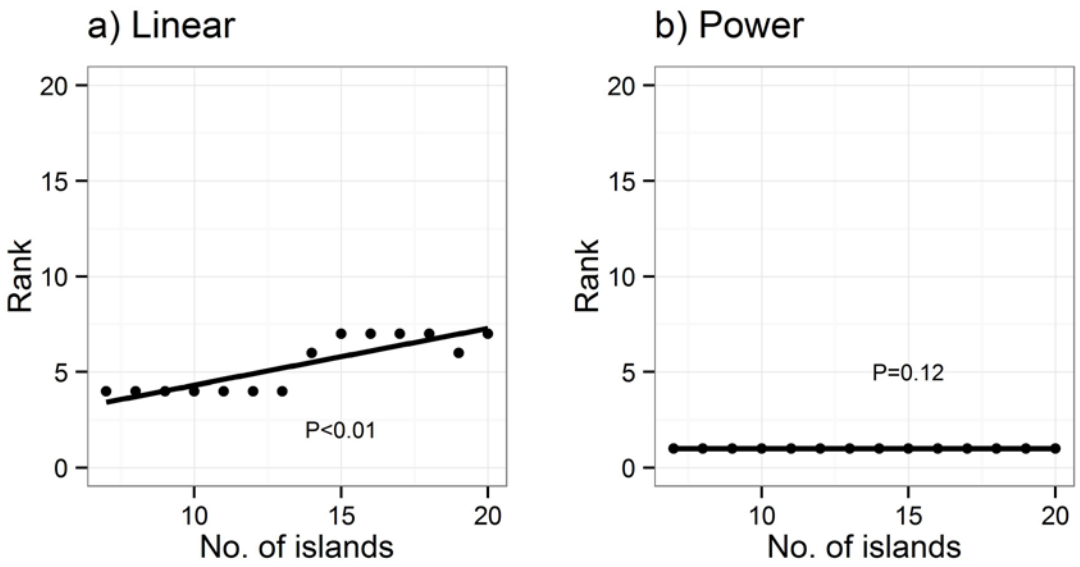

Figure 2 Change in model performance with the number of habitat islands in a dataset. Results for two models are displayed: (a) linear and (b) power models. The rank of a model refers to the model selection analysis in which 20 island species–area relationship models were compared using AIC_c. A model 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.

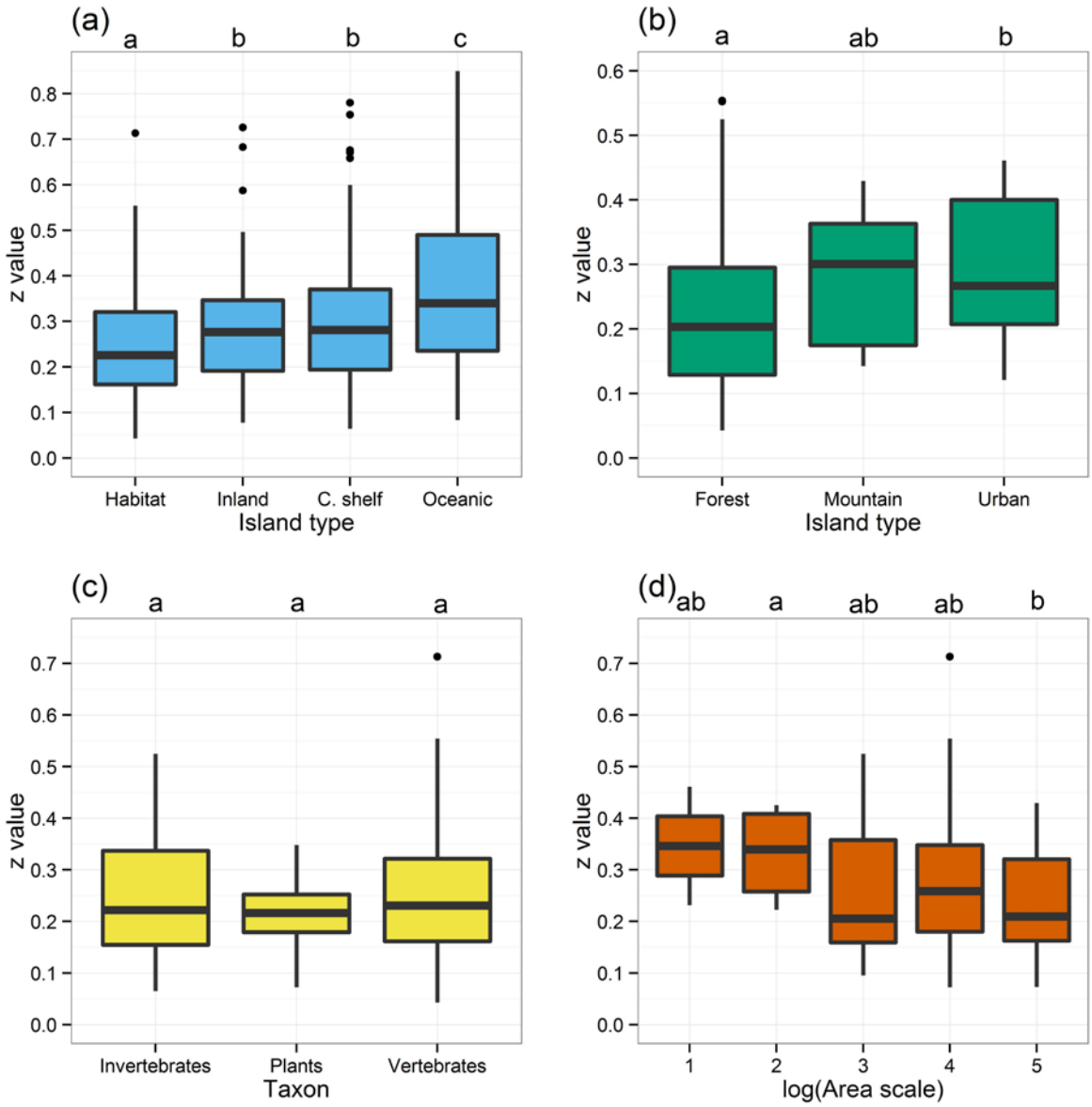

Figure 3 Variation in the *z* parameter of the power (log–log) model across: (a) all island types, (b) different habitat island types, (c) different major taxa, and (d) area scale (log transformed). NB (b)–(d) were calculated using only habitat island datasets. For all plots, only datasets with significant *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.