FORUM: Island, archipelago and taxon effects: mixed models as a means of dealing with the imperfect design of nature's experiments

Nils Bunnefeld* and Albert B. Phillimore

Department of Life Sciences, Imperial College London, Silwood Park, Buckhurst Road, Ascot, SL5 7PY, UK

* Author for correspondence n.bunnefeld06@imperial.ac.uk

A major aim of island biogeography has been to describe general patterns of species richness across islands and to identify the processes responsible. Data are often collected across many islands; with larger datasets providing increased statistical power and more accurate parameter estimates. However, there is often structure in observational data, violating an assumption of linear models that each datum is independent. In island biogeography this structure may take the form of an island, archipelago or taxon being represented by multiple data points. We survey recent papers in this field and find that these forms of non-independence are a common feature. Most authors addressed this problem by conducting separate analyses for each archipelago, taxon or combination of the two, but a better tool for dealing with non-independence and structure in data, the mixed model, already exists. We demonstrate the advantages of a mixed model approach by applying it to a well-known dataset that spans 134 observations of single island endemic (SIE) richness across 39 islands, four archipelagos and four taxa. Taking island area and age into account, SIE richness varies substantially more among archipelagos than it does among islands or taxa. We find that SIE richness rises with island age on the Azores and Galapagos, while on the Canaries and Hawaii SIE richness initially rises with age but later declines on older islands. Our analyses demonstrate three advantages to island biogeography of applying a mixed modelling approach: (1) structure in the data is controlled for; (2) the variance among islands, archipelagos and taxa is estimated; (3) all the data can be included in a single model, making it possible to test whether trends are general across all archipelagos and taxa or are idiosyncratic.

Introduction

A central goal of island biogeography is to understand the processes responsible for generating heterogeneity in biodiversity among islands (MacArthur and Wilson 1963, MacArthur and Wilson 1967, Whittaker and Fernández-Palacios 2007). The most common approach to addressing this question is to compile data on numbers of species per island and then to examine the degree to which species richness is explained by island attributes (particularly area, isolation, habitat heterogeneity and island age) using linear regression techniques. An assumption of linear models is that all data points are independent. However, in island biogeography, as is the case with many strands of non-experimental biology, numerous factors can cause this assumption to be violated (see Table 1 for examples of non-independence in the island biogeography literature). The problem of non-independence of data becomes particularly acute when researchers consider larger datasets in a quest for statistical power and model generality. In this paper, we show that linear mixed modelling (LMM) can address these issues by adding information on the structure of the data (e.g., pseudoreplication due to multiple data points coming from the same island, archipelago or taxon). Moreover, LMMs can offer novel insights by estimating the variation among islands, archipelagos and taxa.

To illustrate the most common forms of non-independence pertinent to island biogeography, we will explore a hypothetical statistical model that aims to address island area as a predictor of the species richness of ten distinct taxa across all of the islands constituting ten different archipelagos. In this case there are at least three sources of non-independence, namely island, archipelago and taxon, which we will consider in turn.

Island Effects: On any single island, a multitude of local factors (including aspects of the environment and island history) not included in our statistical model may make the species richness of the ten different taxa more similar, thereby introducing non-independence (pseudoreplication). For instance, the island may have experienced a recent tropical storm that exterminated many species, resulting in a reduction in the species richness of all ten taxa.

Archipelago Effects: Equally, within a given archipelago, the species richness of the constituent islands may tend to be particularly high or low due to attributes of the archipelago, such as its geological history, climate, inter-island isolation and isolation from sources of colonists. The combined effects of these factors that make species richness more similar across islands within an archipelago is sometimes referred to as *biogeographical coherence* (Santos et al. 2010).

Taxon Effects: Differences in diversity between taxa are a ubiquitous feature of biodiversity and are often associated with intrinsic traits of taxa as well as exogenous environmental factors. For example, just as on the mainland, we may expect a taxon whose representatives are top carnivores to be less species rich across different islands and archipelagos than a taxon made up of herbivores. Within island biogeography, taxon effects on species richness on different islands are likely to be particularly affected by species richness on the mainland and by the degree of inter-island dispersal.

If a single standard linear model was applied to data collected for the hypothetical scenario described above, there would be a high probability of detecting a general trend where there is none (see Fig. 1 for an example of how a false positive or Type I error may arise). A partial solution, often employed in the

study of island biogeography (Table 1), is to construct separate linear models within subsets of data comprising each unique combination of archipelago and taxon identity. The estimated parameters are then compared across the different subsets, with the aim of drawing conclusions about the similarities or differences across the different data subsets. While this approach does remove pseudoreplication due to archipelago and taxon, it leads to very small data sets that offer low power to detect trends, resulting in (i) the failure to find a trend where there is one, also called false negative or Type II error; (ii) uncertainty in parameter estimates; (iii) the linear models are easily overfitted, i.e. too many explanatory variables for the number of data points.

Although analysing each taxon by archipelago combination separately reduces non-independence within tests, any further comparison of models between tests, such as comparing models for different taxa within an archipelago, reintroduces non-independence. If we are testing for a difference in slopes then this pseudoreplication will increase the frequency of Type II errors. Under these conditions, a finding of no significant difference between slopes does not constitute reliable evidence of a shared slope or general trend across taxa. Unfortunately, this latter problem appears to arise quite frequently in island biogeography (Table 1).

Here, we show that linear mixed models (LMMs), a class of models that have gained in popularity in other fields of ecology and evolution (Bolker et al. 2009), can overcome all of these problems and deliver novel insights. Surprisingly, in a survey of recent papers in the field of island biogeography and species richness, we found no case where LMM was used (see Table 1 for details). Two rare instances where a mixed modelling approach (generalised linear mixed models in both cases) has been

used in an island biogeography setting are Blackburn et al. (2004) and Steinbauer et al. (2011).

LMMs, similarly to linear models, describe the degree to which a response variable is predicted by explanatory variables, but the explanatory variables are of two types, fixed effects and random effects (see Glossary for an explanation of LMM terminology). We include the variables for which we intend to estimate slope or intercept coefficients as fixed effects. For instance, if we want to test the hypothesis that species richness increases with island area, we would include area as a fixed effect (just as would be the case in a linear model). Random effects describe the grouping (e.g. taxon) or the hierarchical structure (data points within islands within archipelagos) within the data. Rather than estimating a coefficient for each level of a random effect, we estimate a single parameter, the variance across levels of the random effect. Therefore, random effects provide an efficient means of dealing with non-independence, such as data points from the same island, archipelago or taxon. The number of levels each random effect should have in order to obtain reliable estimates of variance is a matter of debate, with some arguing that more than five are needed (Bolker et al. 2009).

If we return to the hypothetical example from earlier, we can see how this might be treated in a mixed model framework (eq. 1). We are seeking to explain the log transformed species richness +1 (S) of a taxon p on island i within archipelago g as a function of the grand mean (μ) and log island area (A) and its slope coefficient (β). However, we also want to quantify how S of each taxon (p) on each island (i) in each archipelago (g) deviates from the grand mean, and this variance is captured as island (I), archipelago (G) and taxon (P) effects, leaving those differences that are

unique to that particular combination of archipelago, island and taxon (the residual term: *E*).

$$S_{igp} = \mu + \beta A + I_i + G_g + P_p + E_{igp} \qquad \text{eq. 1.}$$

In a mixed modelling framework we can estimate the variance across all levels of *I*, *G*, *P* and *E*. In this way, the variance estimate for island, taxa or archipelago provides us with additional information, whilst also correcting for pseudoreplication. This is commonly called variance components analysis and conveys the amount of variance distributed among the separate random effects (Pinheiro and Bates 2000 p. 50, Zuur et al. 2009 p.).

So far we have considered a LMM with the random effects of island, archipelago, taxon and residual. This model type is also called a random intercept model because island, archipelago and taxon are categorical variables with a number of levels across which we estimate variation around the grand mean. Consider a more complex case where we are interested not only in whether archipelagos vary in their intercept but also in whether the slope of the island area and species richness relationship varies among archipelagos. We can extend the LMM approach to a random slope model (sometimes called random regression), where, we can estimate the variation of the slope across different archipelagos (or taxa). Conceptually, the random slope (or random regression) model is similar to an analysis of covariance (ANCOVA), where we estimate different slopes for each level of a categorical variable by adding the interaction between the continuous variable and the categorical variable. The difference being that the object of interest is the variance in slopes rather than the coefficient of each slope.

We illustrate the application of a mixed modelling approach to data on the richness of single island endemics (SIEs) collected for multiple islands, archipelagos and taxa. These data were originally used to test the general dynamic model (GDM) of oceanic island biogeography (Whittaker et al. 2008, 2010). The GDM posits that the number of SIEs initially rises and later falls through time, coincident with changes in island area, altitude and habitat diversity (see also Stuessy 2007, Whittaker et al. 2007). Whittaker et al. (2008, 2010) applied separate linear models to unique combinations of archipelago and taxon and showed that by including area, time and time^2 as predictors (their ATT^2 model) they could explain a large proportion of the variance in SIE richness. Note that in the ATT^2 model the rise and fall of SIE richness is independent of island area. In this study we test whether the influences of area, time and time^2 on SIE richness predicted by the GDM persist when a mixed modelling approach is applied and additional variables are considered. We also highlight the further insights that a mixed modelling approach provides.

Methods

The island data set

Data (n=134) on the number of single island endemics (SIEs), island age (in millions of years) and area (km^2) were kindly provided by R. J. Whittaker and K. A. Triantis (Whittaker et al. 2008, Whittaker et al. 2010). Data from four volcanic archipelagos, the Azores (Arthropods, Coleoptera and land snails), Canaries (Arthropods, Coleoptera, land snails and plants), Galapagos (Arthropods, Coleoptera, and plants) and Hawaii (Arthropods, Coleoptera, land snails and plants) were included. See

Whittaker et al. (2008) for a full discussion of the selection of islands and for basic evaluation of data quality issues and important assumptions regarding age of islands. The data analysed here on Arthropods and Coleoptera on the Azores (Borges et al. 2005, Borges and Hortal 2009) and Coleoptera on the Canaries (Izquierdo et al. 2004) were not included in the study of Whittaker et al. (2008). Data on isolation from the nearest island and from the mainland (both in km) were obtained from UNEP island directory (http://islands.unep.ch/isldir.htm) and Silva & Smith (2004).

Statistical analyses

Our response variable was the number of single island endemics (SIE), a diversity metric that has been used by a number of recent studies. For the sake of simplicity we apply a ln(*n*+1) transformation to the response variable, meaning that linear mixed modelling techniques can be applied. However, we refer the reader to a recent study demonstrating that when faced with count data of this sort, a negative binomial or quasi-Poisson error structure often performs better than ln(n+1) transformation (O'Hara and Kotze 2010). Our fixed effects were the area of the island (km^2), distance to the mainland (km), distance to the nearest island (km), the geological age of the island (in millions of years) and its quadratic term.

Following Zuur et al. (2009), we first selected the most parsimonious random intercepts structure by finding the model with the lowest Akaike Information Criterion corrected for small sample size (AICc) with all fixed effects added. ΔAICc is calculated as the difference between each model's AICc and the lowest AICc, with a ΔAICc < 2 interpreted as substantial support that the model belongs to the set of

best models, a ΔAICc of 4 – 7 corresponding to less support and ΔAICc > 10 treated as providing no support that the model belongs to the best set. Akaike weights give the probability that a model is the best model, given the data and the set of candidate models (Burnham and Anderson 2004).

Models were fit with lmer in the lme4 package (version 0.999375-37) in R (R Development Core Team 2010). When comparing models that varied in their random effects but not fixed effects, the models were fit using restricted maximum likelihood (REML). To find the most parsimonious random intercept structure we ran models with archipelago, island, taxon and taxon:archipelago (meaning the unique combination of taxon and archipelago) as random effects. Island is nested within archipelago because each island occurs only in one archipelago. Taxon is crossed random effect with respect to both island and archipelago, since each taxon can occur on more than one island or archipelago (see Glossary for definition of nested versus crossed random effects). For simplicity we assume that the taxa are independent of each other. In reality this assumption is likely to be violated because Coleoptera are a subclade of Arthropods and plants are more distantly related to snails, beetles and Arthropods than the latter are to each other. Methods for dealing with phylogenetic non-independence in a mixed model setting do exist (Hadfield and Nakagawa 2010), however, we will not consider this issue further here. Equally, pseudoreplication arising via island or archipelago effects can be seen as points along a continuum of spatial autocorrelation. We draw the reader's attention to the option of dealing with this continuum directly in a mixed modelling framework by incorporating spatial variation and covariation in the error term of the mixed model

(note that this is not possible in the lme4 library but can be done using the nlme library).

Whittaker et al. (2008) argue that the curve of the rise and fall of diversity with island age will vary among archipelagos. Therefore, we use a random slope model to test whether allowing the age versus SIE richness slope to vary among archipelago:taxon combinations improved model performance.

After determining the random effect structure, the most parsimonious combination of fixed effects must be found using maximum likelihood (ML) rather than REML. We conducted AICc based multi-model inference using the dredge function in the MuMIn package in R (version 0.13.17) to run a complete set of models with all possible combinations of the fixed effects. The code in R for all mixed model analyses is given in Appendix 1.

Graphical inspection is an important tool in statistics (Hilborn and Mangel 1997). After graphical inspection of the final model (Pinheiro and Bates 2000, Zuur et al. 2009), major differences in dynamics among archipelagos were apparent, and on this basis we chose to explore the alternative analysis of adding archipelago as a fixed effect instead of as random effect (see below). Distance to mainland was not included as fixed effect in these models due to low variation within archipelagos.

Results

Random effects

The lowest AICc random effects structure included archipelago, island, taxon and taxon:archipelago as random intercepts (Table 2). Fitting either a random slope of area or age for each taxon:archipelago led to an increase in AICc (i.e., the model was poorer) (Table 2). In the most parsimonious (lowest AICc) model, after fitting the fixed effects (age of the island, squared age of the island and the area of the island), archipelago explained 59% of the variation in the random effects, with the remainder being distributed as follows island = 7%, taxon:archipelago = 3%, taxon = 23% and residual = 8%.

Fixed effects

The lowest AICc model shows the number of SIEs increasing with both the age and area of the island but with a negative quadratic slope for age (Table 3, Appendix 2): species richness increases with area but rises and then falls over time. This model has nine estimated parameters, four fixed effects and five random effects. Neither distance to the mainland and distance to the nearest island improved model fit (Table 3). Inspection of predicted versus observed values showed that the model explained most of the variation in SIE richness (Appendix 3).

An alternative approach

As archipelago effects were much more sizeable than any of the other random effects and due to the substantial variation in maximum island age among archipelagos, we investigated whether the slopes of the effects of island age or area differed among archipelagos. While there was no support for these hypotheses when we used a random slope approach (Table 2), we were concerned that this

might be due to the small number of archipelagos across which we were trying to estimate variance in both intercept and slope. We therefore adopted an alternative approach and included archipelago as a fixed effect and tested whether allowing for an interaction between archipelago and various covariates (island age, age^2 and area) led to an improved model fit. By fitting archipelago as a fixed effect and allowing for interaction with a covariate, we estimated both an intercept and a slope for each archipelago.

Treating archipelago as a fixed effect led to a substantial improvement in model fit (ΔAICc = 11.1, models estimated with maximum likelihood). The most parsimonious model included archipelago, age of the island, squared age of the island and the area of the island (Figure 2) as fixed effects. It also included the interaction of archipelago with squared age of the island as fixed effect (Table 4, Figure 3). The variance of the random effects included were as follows: island: 12%, taxon:archipelago: 7%, taxon: 56%, residual: 25% (Appendix 2). This model is based on 14 estimated parameters, 10 fixed effects and four random effects. Graphical inspection showed that the model explained most of the variation in SIE richness (Appendix 3). Across all four archipelagos this model described an increase in SIE richness with island area (Figure 2). However, the effect of island age differed among archipelagos. A positive near linear increase in SIE richness with island age was observed on the youngest two archipelagos, the Azores and Galapagos. In comparison, the coefficients estimated for the Canaries and Hawaii described a rise and fall of SIE richness with island age.

Discussion

By applying a mixed model approach to island SIE richness data across four volcanic oceanic archipelagos, we found that the ATT^2 formulation proposed by Whittaker et al. (2007, 2008, 2010) provides an excellent description of the data. We also identified previously unappreciated nuances that we will discuss below. Plotting the predicted $T+T^2$ function for the different archipelagos (Fig. 2) revealed that on the Canaries and Hawaii SIE richness rose and fell with increasing island age, exactly as predicted by Whittaker and colleagues. In both cases islands showed an increase in diversity for the first 10MY and a decrease for the next 10MY. In comparison and consistent with Borges and Hortal (2009), we found that on the Azores and Galapagos SIE richness only increased with island age (Fig. 2).

The increase in SIE richness with island age on younger islands has a intuitive explanation; speciation, whether by anagenesis or cladogenesis, requires time (Price 2008, Rosindell and Phillimore 2010). Moreover, early in an island's history, the more species present, the more species that are available to undergo renewed bouts of cladogenesis (Whittaker et al. 2007). A slowing of the rate of SIE accumulation with time (not evident in the Azores) makes sense in the light of accumulating phylogenetic and fossil evidence that there are limits to diversity and that speciation rates are regulated by diversity (Phillimore and Price 2008, Rabosky and Glor 2010, Ezard et al. 2011). However, an explanation for the decrease in the SIE richness of the older islands is less straightforward. In describing the GDM, Whittaker et al. (2008, 2010) suggested that SIE richness on old islands may fall as a consequence of

a reduction in island area (we can discount this explanation as area is included in the ATT^2 model), topographic complexity and habitat diversity or by SIEs colonizing other islands and becoming multiple island endemics rather than SIEs. An alternative explanation is the taxon cycle model as described by Ricklefs and colleagues (Ricklefs and Cox 1972, Ricklefs and Bermingham 2002); perhaps old endemic species on the oldest islands are most susceptible to immigration of new colonists and the diseases that they carry.

The mixed modelling approach provided insights into how archipelagos, taxa and islands vary in SIE richness. For instance, by far the most substantial variance component was the archipelago effect, with taxon effect also substantial. This suggests that biogeographical coherence has a large influence on SIE richness. We found evidence for a significant taxon:archipelago interaction, meaning that the SIE richness of a particular taxon on a particular island was not entirely accounted for by the main random effects, although this interaction captured relatively little variance. Island effects and the residual term were both small.

Whittaker et al. (2008) fitted 14 different models, one for each island-taxon combination, to explain the number of single island endemics using area, the time and the squared time as explanatory variables. Thus, for each of their models, they estimated four parameters; three slopes and one intercept, summing to 56 parameters. In our lowest AICc model, treating archipelago as fixed effect, we used 14 parameters in total, 10 fixed effects, and four random effects. The advantages of the mixed modelling framework over separate fitting of terms to each archipelago:taxon combination are parsimony and generality, since we have used

fewer parameters in a single model framework, whilst modelling differences among islands, archipelagos and taxa.

A challenge in the application of mixed models to data is their added complexity. For example, they require the user to have thought about which variables to include as fixed effects and random effects and the nesting structure of random effects. Secondly, even though mixed models can deal with unbalanced design (different number of data points for different groups of the random effect), LMMs require a data set that is large enough to estimate variances for each group of the random effects. Here we have chosen to exemplify the advantages of using linear mixed models for island biogeography using the lme4 package in the free software R. Within R, there are several packages for LMMs and many other common software packages can fit LMMs (see Bolker et al. 2009). The choice of the package depends on the data and the research question. LMMs are a flexible tool and recent developments have extended linear mixed models to generalized linear mixed models (GLMM) with the option to include a link function and error structures to allow for the non-normal distribution of the response variable such as species richness data (often Poisson distributed) or binomial data for zero-one-states such as alive or dead (Bolker et al. 2009). Further reading on LMMs and GLMMs and statistical detail can be obtained from Pinheiro and Bates (2000), Zuur et al. (2009) and Hadfield (2010).

In summary, linear mixed models increase the power to detect general patterns where data come from grouped sources, such as is common in island biogeography and have the potential to offer additional insights that linear models

cannot. Therefore, we suggest that this approach should be adopted as standard in future island biogeography studies that have the goal of identifying general trends across archipelagos and taxa.

Acknowledgments

We are very grateful to K. A. Triantis for providing data and T. J. Ezard, D. Fontaneto, J. Hortal, S. Meiri, D. Orme, T. Price, J. Rosindell, K. A. Triantis, R. J. Whittaker and three anonymous reviewers for comments that greatly improved the manuscript. NB is supported by the European Commission under the HUNT project of the 7th Framework Programme for Research and Technological Development. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use made of the information. The views expressed in this publication are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission. ABP is funded by an Imperial College Junior Research Fellowship.

References

Blackburn, T. M., et al. 2004. Avian extinction and mammalian introductions on oceanic islands. - Science 305: 1955-1958.

Bolker, B., et al. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. - Trends Ecol Evol 24: 127-135.

Borges, P. A. V., et al. 2005. A list of the terrestrial fauna (*Mollusca* and *Arthropoda*) and flora (*Bryophyta*, *Pteridophyta* and *Spermophyta*) from the Azores. - In: Direcção Regional do Ambiente and Universidade dos Açores, Horta, Angra do Heroísmo and Ponta Delgada.

Borges, P. A. V. and Hortal, J. 2009. Time, area and isolation: factors driving the diversification of Azorean arthropods. - Journal of Biogeography 36: 178-191.

Burnham, K. P. and Anderson, D. R. 2004. Multimodel inference: Understanding AIC and BIC in model selection. - Sociological Methods and Research 33: 261-304.

Cardoso, P., et al. 2010. Drivers of diversity in Macaronesian spiders and the role of species extinctions. - Journal of Biogeography 37: 1034-1046.

Dengler, J. 2010. Robust methods for detecting a small island effect. - Divers. Distrib. 16: 256-266.

Dexter, K. G. 2010. The influence of dispersal on macroecological patterns of Lesser Antillean birds. - Journal of Biogeography 37: 2137-2147.

Ezard, T. H. G., et al. 2011. Interplay between changing climate and species' ecology drives macroevolutionary dynamics. . - Science 332: 349-351.

Fattorini, S. 2009. On the general dynamic model of oceanic island biogeography. - Journal of Biogeography 36: 1100-1110.

Hadfield, J. D. 2010. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. - Journal of Statistical Software 33: 1-22.

Hadfield, J. D. and Nakagawa, S. 2010. General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. - Journal of Evolutionary Biology 23: 494-508.

Hannus, J. J. and von Numers, M. 2010. Temporal changes in the island flora at different scales in the archipelago of SW Finland. - Appl. Veg. Sci. 13: 531-545.

Hilborn, R. and Mangel, M. 1997. The Ecological Detective: Confronting models with Data. - Princeton University Press.

Hortal, J., et al. 2009. Island Species Richness Increases with Habitat Diversity. - American Naturalist 174: E205-E217.

Horvath, R., et al. 2009. Spiders are not less diverse in small and isolated grasslands, but less diverse in overgrazed grasslands: A field study (East Hungary, Nyirseg). - Agric. Ecosyst. Environ. 130: 16-22.

Izquierdo, I., et al. 2004. Lista de Especies Silvestres de Canarias (Hongos, Plantas y Animales Terrestres). - In: Consejería de Política Territorial y Medio Ambiente, Gobierno de Canarias, La Laguna, Tenerife.

Jonsson, M., et al. 2009. Patterns of invertebrate density and taxonomic richness across gradients of area, isolation, and vegetation diversity in a lake-island system. - Ecography 32: 963-972.

Kallimanis, A. S., et al. 2010. Biogeographical determinants for total and endemic species richness in a continental archipelago. - Biodivers Conserv 19: 1225-1235.

Keppel, G., et al. 2010. Drivers of lowland rain forest community assembly, species diversity and forest structure on islands in the tropical South Pacific. - J. Ecol. 98: 87-95.

Long, J. D., et al. 2009. Linking invasions and biogeography: Isolation differentially affects exotic and native plant diversity. - Ecology 90: 863-868.

MacArthur, R. H. and Wilson, E. O. 1963. An equilibrium theory of insular zoogeography. - Evolution 17: 373-387.

MacArthur, R. H. and Wilson, E. O. 1967. The Theory of Island Biogeography. - Princeton University Press.

Murakami, M. and Hirao, T. 2010. Lizard predation alters the effect of habitat area on the species richness of insect assemblages on Bahamian isles. - Divers. Distrib. 16: 952-958.

O'Hara, R. B. and Kotze, D. J. 2010. Do not log-transform count data. - Methods in Ecology & Evolution 1: 118–122.

Phillimore, A. B. and Price, T. D. 2008. Density-dependent cladogenesis in birds. - PLoS Biology 6: e71.

Pinheiro, J. C. and Bates, D. M. 2000. Mixed Effects Models in S and S-Plus. - Springer.

Price, T. D. 2008. Speciation in birds. - Roberts & Company Publishers.

R Development Core Team 2010. R: A language and environment for statistical computing. - In: R Foundation for Statistical Computing. URL http://www.R-project.org.

Rabosky, D. L. and Glor, R. E. 2010. Equilibrium speciation dynamics in a model adaptive radiation of island lizards. - Proc. Nat. Acad. Sci. USA 107: 22178-22183.

Ricklefs, R. E. and Cox, G. W. 1972. Taxon cycles in the West Indian avifauna. - American Naturalist 106: 195-219.

Ricklefs, R. E. and Bermingham, E. 2002. The concept of the taxon cycle in biogeography. - Glob. Ecol. Biogeogr. 11: 353-361.

Rosindell, J. and Phillimore, A. B. 2010. A unified model of island biogeography sheds light on the zone of radiation. - Ecology Letters 14: 552–560.

Santos, A. M. C., et al. 2010. Are species-area relationships from entire archipelagos congruent with those of their constituent islands? - Glob. Ecol. Biogeogr. 19: 527-540.

Silva, L. and Smith, C. W. 2004. A characterization of the non-indigenous flora of the Azores archipleago. - Biological Invasions 6: 193-204.

Steinbauer, M. J. and Beierkuhnlein, C. 2010. Characteristic pattern of species diversity on the Canary islands. - Erdkunde 64: 57-71.

Steinbauer, M. J., et al. 2011. Increase of island endemism with altitude – speciation processes on oceanic islands. - Ecography In Press:

Stracey, C. M. and Pimm, S. L. 2009. Testing island biogeography theory with visitation rates of birds to British islands. - Journal of Biogeography 36: 1532-1539.

Stuessy, T. F. 2007. Evolution of specific and genetic diversity during ontogeny of island floras: the importance of understanding process for interpreting island biogeographic

patterns. - In: M. C. Ebach and R. S. Tangney (eds), Biogeography in a Changing World. CRC Press.

Tuya, F. and Haroun, R. J. 2009. Phytogeography of Lusitanian Macaronesia: biogeographic affinities in species richness and assemblage composition. - Eur. J. Phycol. 44: 405-413.

Uchida, Y. and Inoue, M. 2010. Fish species richness in spring-fed ponds: effects of habitat size versus isolation in temporally variable environments. - Freshw. Biol. 55: 983-994.

Whittaker, R. J. and Fernández-Palacios, J. M. 2007. Island Biogeography: Ecology, Evolution and Conservation. - Oxford University Press.

Whittaker, R. J., et al. 2007. The island immaturity - speciation pulse model of island evolution: an alternative to the "diversity begets diversity" model. - Ecography 30: 321-327.

Whittaker, R. J., et al. 2008. A general dynamic theory of oceanic island biogeography. - Journal of Biogeography 35: 977-994.

Whittaker, R. J., et al. 2010. A general dynamic theory of oceanic island biogeography: extending the MacArthur–Wilson theory to accommodate the rise and fall of volcanic islands. - In: J. B. Losos and R. E. Ricklefs (eds), The Theory of Island Biogeography Revisited. Princeton University Press, pp. 88-115.

Zuur, A. F., et al. 2009. Mixed effects models and extensions in ecology with R. - Springer.

GLOSSARY

Crossed random effect – multiple random effects that apply independently to a data point such as a species belongs to a taxon and was found on a specific island where both island and taxon influence the occurrence of the species.

Fixed effect – estimates intercepts for factor levels or slopes for continuous variable. Examples are age or area of islands.

Mixed effects model – a model with both fixed and random effects.

Nested random effects – multiple random effects with a hierarchical structure where levels of one random effect may only vary within a level of another random effect. For instance, island effects are nested within archipelago effects.

Random effects – grouping units drawn as a random sample from a population, for example islands, archipelagos or taxa. Allow estimating the variance explained by the grouping unit, such as different islands.

Random intercept model – a model where the deviations of the intercepts of the levels of the random effect from the grand mean come from a normal distribution with an estimated variance.

Random slope/regression model – extending the random intercept model by allowing the slope estimated for a fixed effect to vary among the individual units of a random effect. For example, the slope of the relationship between age of the island and species richness varies on different islands.

Table 1. Summary of the forms of pseudoreplication that exist in recent island biogeography studies, the approaches taken to remedy it and suggestions for how linear mixed models could improve on this.

Study	Type of study[†]	Potential pseudoreplication[‡]	Method to deal with pseudoreplication employed in paper	Suggested model structure in LMM[¥]
Dexter 2010	Species area relationship	Island, habitat	Separate analysis for each habitat	Fixed = habitat Random = island
Murakami and Hirao 2010	Species area relationship	Island, taxon	Separate analysis for each taxon	Random = island + taxon
Hannus and von Numers 2010	Species area relationship	Island, year	Separate analysis for each year	Fixed = year Random = island
Santos et al. 2010	Species area relationship	Island, archipelago, taxon	Separate analysis for each combination of archipelago and taxon	Random = island + archipelago + island + archipelago:taxon
Steinbauer and Beierkuhnlein 2010	Correlates of species richness	Island, taxon	Separate analysis for each taxon	Random = island + taxon
Cardoso et al. 2010	Correlates of species richness	Archipelago	Separate analysis for each archipelago	Fixed = archipelago
Uchida and Inoue 2010	Correlates of species richness	Island, sampling time	Mean taken across sampling times.	Random = island + sampling time
Kallimanis et al. 2010	Correlates of species richness	Biogeographic region within archipelago	No action taken	Random = biogeographic region within archipelago

Dengler 2010	Species area relationship	Floristic region within archipelago	No action taken	Random = floristic region within archipelago
Ishtiaq, et al. 2010	Species area relationship	Island	Separate analysis for each taxon	Fixed = taxon Random = island
Keppel et al. 2010	Correlates of species richness	Archipelago	No action taken	Random = archipelago
Jonsson et al. 2009	Species area relationship	Island	Separate analysis for each combination of taxon and year	Fixed = taxon + year Random = island
Hortal et al. 2009	Habitat area relationship	Island, archipelago, taxon	Separate analysis for each combination of archipelago and taxon	Random = island + archipelago + island + archipelago:taxon
Tuya and Haroun 2009	Correlates of species richness	Island, archipelago, taxon	Separate analysis for each archipelago	Fixed = archipelago + taxon Random = island
Stracey and Pimm 2009	Correlates of species richness	Island, species type (visitor/resident)	Separate analysis for visitors and residents	Fixed = species type Random = island
Fattorini 2009	Correlates of species richness	Island, taxon	Separate analysis for each taxon	Random = island + taxon
Long et al. 2009	Correlates of species richness	Island, species type (native/exotic)	Separate analysis for native and exotic species	Fixed = species type Random = island
Horvath et al. 2009	Correlates of species richness	Fragment, habitat	Separate analysis for each habitat	Fixed = habitat Random = island

[†]Studies of species richness include those investigating subsets of species richness, e.g., numbers of single island endemics.

‡ We searched Thomson ISI Web of Knowledge (http://apps.isiknowledge.com/) for papers published over the period 2009-2010 with the keyword “island biogeography”. We only include papers in this table that were addressing correlates of species richness across islands or patches of some sort and where the assumption of independence of datapoints appeared to be violated.

¥We suggest a model structure that would remove the pseudoreplication in a mixed modelling framework. We only include variables as random effects where there are more than four levels, otherwise we propose treating the variable as a fixed effect (Bolker et al. 2009). In many cases it would also be appropriate to fit interactions between the fixed effects for archipelago (or taxon) and other fixed effects, such as area. Alternatively, such interactions could be fitted as random effects and random regression implemented.

Table 2: AICc for random effects model selection with single island endemics as response variable and distance to mainland, age, age^2, area and distance to nearest island as fixed effects (n=134). Random intercepts were included for archipelago, island, taxon and taxon:archipelago. Random slopes for area and age varying across taxon:archipelago combinations were tested. An X indicates that a variable was included in the model, whereas a blank field means that the variable was not included.

Random intercept				Random slope		
Archipelago	Island	Taxon	Taxon:archipelago	Area within taxon: archipelago	Age within taxon: archipelago	AICc
X	X	X	X			330.80
X	X	X	X		X	334.24
X	X	X	X	X		335.26
X	X		X			338.71
X	X	X				341.09
	X	X	X			342.95
X		X	X			350.91

Table 3: Coefficients for the fixed effects of the four most parsimonious models that treat archipelago, island, taxon and taxon:archipelago as random effects. The number of parameters in the model (k), the AICc difference (Δ AICc) and AICc weight is given for each model.

Intercept	Distance mainland	Age	Age^2	Area	Nearest island	k	AICc	Δ AICc	AICc weight
-0.3117		0.2410	-0.0129	0.4466		9	282.6	0.00	0.51
-0.4606		0.2336	-0.0129	0.4639	0.0018	10	284.3	1.7	0.22
-0.7650	0.0003	0.2518	-0.0132	0.4535		10	284.6	2.0	0.19
-0.8565	0.0002	0.2433	-0.0131	0.4693	0.0017	11	286.4	3.8	0.08

Table 4: The fixed effects included in the three most parsimonious models that treat archipelago as a fixed effect. The number of parameters in the model (k), the AICc difference (Δ AICc) and AICc weight is given for each model. The random effects for the models were island, taxon and taxon:archipelago. Archipelago (ARCH) was included as a fixed effect and its interactions with age, age^2 and area. Distance to mainland was not included as fixed effect due to low variation within archipelagos. An X indicates that a variable was included in the model, whereas a blank field means that the variable was not included.

ARCH	Age	Age^2	Area	Nearest island	ARCH: age	ARCH: age^2	ARCH: area	k	AICc	ΔAICc	AICc weight
X	X	X	X			X		14	271.5	0.0	0.24
X	X	X	X					11	272.2	0.25	0.17
X	X	X	X		X	X		17	272.9	1.43	0.12

Figure 1. A hypothetical scenario illustrating how archipelago or taxon effects can mislead. In this figure there is no underlying relationship (dotted line) between island area and species richness but data points from the same archipelago (or taxon) tend to be more similar than those from different archipelagos (or taxa). In this case the slope estimated using a linear model (grey solid line) incorrectly identifies a positive relationship across all islands, which is due to differences between archipelagos.

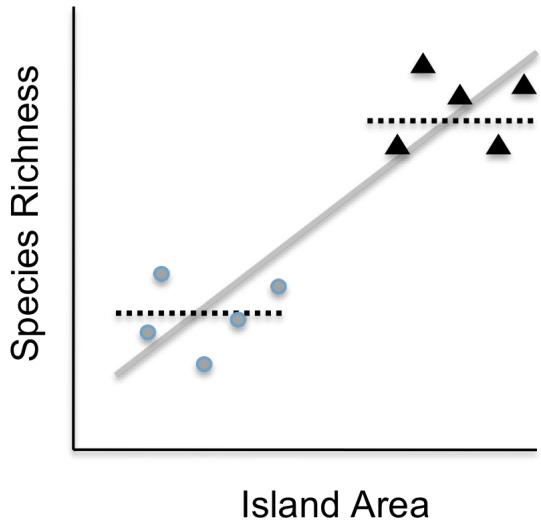
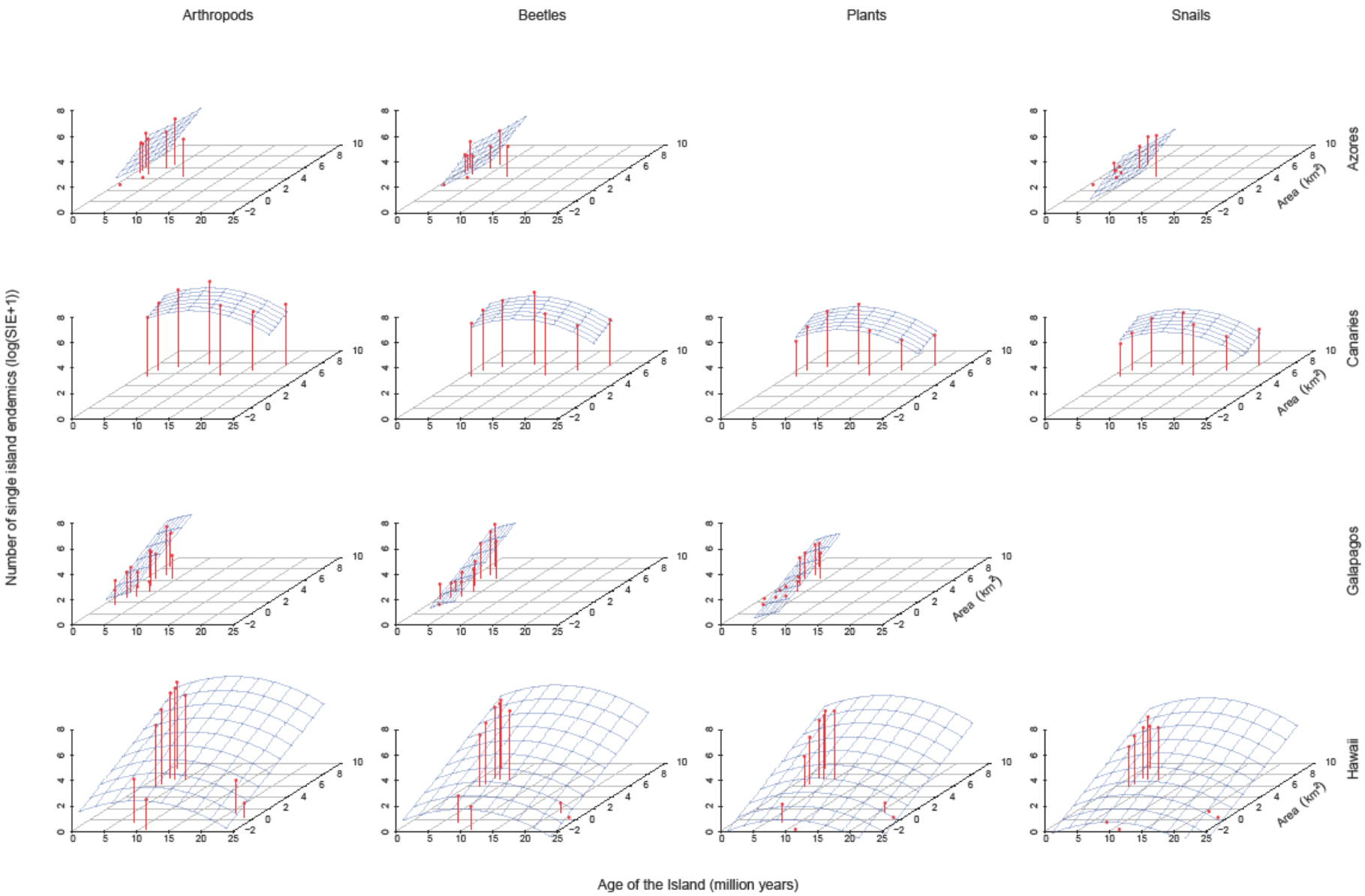


Figure 2: Prediction from the lowest AICc mixed effect model including the random effects of taxon:archipelago, island and taxon. The response variable SIE was ln(n+1) transformed and explained by area of the island and age of the island (n=134). The grid is predicted from the fixed effects estimates (termed best linear unbiased estimates) from the model with the lowest AICc (Table 4) adjusted by the best linear unbiased predictor of the random effect of taxon:archipelago. Please note that the lines are an approximation and do not include the differences due to the random effect of Island. **SEE ALSO SEPERATE FILE FOR FIGURE 2 in higher resolution**


Arthropods
Beetles
Plants
Snails
Azores
Canaries
Galapagos
Hawaii
Number of single island endemics (log(SIE+1))
Age of the Island (million years)
Area (km²)