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

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.

A central goal of island biogeography is to understand the processes responsible for generating heterogeneity in biodiversity among islands (MacArthur and Wilson 1963, 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 1) the failure to find a trend where there is one, also

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.

<table>
<thead>
<tr>
<th>Study</th>
<th>Type of study</th>
<th>Potential pseudoreplication</th>
<th>Method to deal with pseudoreplication employed in paper</th>
<th>Suggested model structure in LMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dexter 2010</td>
<td>Species area relationship</td>
<td>Island, habitat</td>
<td>Separate analysis for each habitat</td>
<td>Fixed = habitat</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Random = island</td>
</tr>
<tr>
<td>Murakami and Hirao 2010</td>
<td>Species area relationship</td>
<td>Island, taxon</td>
<td>Separate analysis for each taxon</td>
<td>Fixed = island + taxon</td>
</tr>
<tr>
<td>Hannus and von Numer 2010</td>
<td>Species area relationship</td>
<td>Island, year</td>
<td>Separate analysis for each year</td>
<td>Random = island</td>
</tr>
<tr>
<td>Santos et al. 2010</td>
<td>Species area relationship</td>
<td>Island, archipelago, taxon</td>
<td>Separate analysis for each combination of archipelago and taxon</td>
<td>Random = island + archipelago + island + archipelago:taxon</td>
</tr>
<tr>
<td>Steinbauer and Beierkuhnlein 2010</td>
<td>Correlates of species richness</td>
<td>Island, taxon</td>
<td>Separate analysis for each taxon</td>
<td>Random = island + taxon</td>
</tr>
<tr>
<td>Cardoso et al. 2010</td>
<td>Correlates of species richness</td>
<td>Archipelago</td>
<td>Separate analysis for each archipelago</td>
<td>Fixed = archipelago</td>
</tr>
<tr>
<td>Uchida and Inoue 2010</td>
<td>Correlates of species richness</td>
<td>Island, sampling time</td>
<td>Mean taken across sampling times</td>
<td>Random = island + sampling time</td>
</tr>
<tr>
<td>Kallimanis et al. 2010</td>
<td>Correlates of species richness</td>
<td>Biogeographic region within archipelago</td>
<td>No action taken</td>
<td>Random = biogeographic region within archipelago</td>
</tr>
<tr>
<td>Dengler 2010</td>
<td>Species area relationship</td>
<td>Floristic region within archipelago</td>
<td>No action taken</td>
<td>Random = floristic region within archipelago</td>
</tr>
<tr>
<td>Ishiaq et al. 2010</td>
<td>Species area relationship</td>
<td>Island</td>
<td>Separate analysis for each taxon</td>
<td>Fixed = taxon</td>
</tr>
<tr>
<td>Keppel et al. 2010</td>
<td>Correlates of species richness</td>
<td>Archipelago</td>
<td>No action taken</td>
<td>Random = archipelago</td>
</tr>
<tr>
<td>Jonsson et al. 2009</td>
<td>Species area relationship</td>
<td>Island</td>
<td>Separate analysis for each combination of archipelago and taxon</td>
<td>Random = island + archipelago + island + archipelago:taxon</td>
</tr>
<tr>
<td>Hortal et al. 2009</td>
<td>Habitat area relationship</td>
<td>Island, archipelago, taxon</td>
<td>Separate analysis for each combination of archipelago and taxon</td>
<td>Random = island + archipelago + island + archipelago:taxon</td>
</tr>
<tr>
<td>Tuya and Haroun 2009</td>
<td>Correlates of species richness</td>
<td>Island, archipelago, taxon</td>
<td>Separate analysis for each archipelago</td>
<td>Fixed = archipelago + taxon</td>
</tr>
<tr>
<td></td>
<td>Correlates of species richness</td>
<td>Island, species type (visitor/resident)</td>
<td>Separate analysis for each archipelago</td>
<td>Random = island</td>
</tr>
<tr>
<td></td>
<td>Correlates of species richness</td>
<td>Island, species type (native/exotic)</td>
<td>Separate analysis for each archipelago</td>
<td>Fixed = species type</td>
</tr>
<tr>
<td></td>
<td>Correlates of species richness</td>
<td>Island, species type (native/exotic)</td>
<td>Separate analysis for each archipelago</td>
<td>Random = island</td>
</tr>
<tr>
<td></td>
<td>Correlates of species richness</td>
<td>Fragment, habitat</td>
<td>Separate analysis for each habitat</td>
<td>Fixed = habitat</td>
</tr>
</tbody>
</table>

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.
are testing for a difference in slopes then this pseudoreplication non-independence due to pseudoreplication of islands. 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 ln(species richness + 1), denoted $S$, for observation $r$ of a taxon $p$ on island $i$ in archipelago $g$ as a function of the grand mean ($\mu$) and log island area ($A$) and its slope coefficient ($b$). We also want to quantify the amount by which the observed $S$ for a particular island:taxon combination $r$ deviates from the best linear unbiased estimate obtained from the fixed effects. In equation 1, $P$ is the deviation for taxon $p$, $G$ is the deviation for archipelago $g$, $I_1$ is the deviation from the mean prediction for island $i$ within archipelago $g$, and $\epsilon_{pgi}$ is the deviation of the observation $r$ from the mean prediction for taxon $p$, archipelago $g$ and island $i$ (known as the residual or error term).

$$S_{pgi} = \mu + \beta A_{pgi} + P + G + I_1 + \epsilon_{pgi}$$  

(1)

In a mixed modelling framework we can estimate the variance across all levels of $P$, $G$, $I$ and $\epsilon$ around a mean of zero. In this way, the variance estimate across taxa, archipelagos and islands 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).

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 taxae. 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 (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 time² as predictors (their ATT² model) they could explain a large proportion of the variance in SIE richness. Note that in the ATT² 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 time² 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 (km²) were kindly provided by R. J. Whittaker and K. A. Triantis (Whittaker et al. 2008, 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 and 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 (km²), 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 Akaikes 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. Akaikes 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 (ver. 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 package but can be done using the nlme package).

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 (ver. 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 Supplementary material 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, Supplementary material 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 (Supplementary material 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, age² 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 (Fig. 2) as fixed effects. It also included the interaction of archipelago with squared age of the island as fixed effect (Table 4, Supplementary material Appendix 3). The variance of the random effects included were as follows: island: 12%, taxon:archipelago: 7%, taxon: 56%, residual: 25% (Supplementary material 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 (Supplementary material Appendix 3). Across all four archipelagos this model described an increase in SIE richness with island area (Fig. 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 ATT² 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 + T² 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 10 my and a decrease for the next 10 my. 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

---

Table 2. AICc for random effects model selection with single island endemics as response variable and distance to mainland, age, age², 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.

<table>
<thead>
<tr>
<th>Archipelago</th>
<th>Island</th>
<th>Taxon</th>
<th>Taxon:archipelago</th>
<th>Area within taxon:archipelago</th>
<th>Age within taxon:archipelago</th>
<th>AICc</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>330.80</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>334.24</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>335.26</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>338.71</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>341.09</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>342.95</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>350.91</td>
</tr>
</tbody>
</table>
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 (ΔAICc) and AICc weight is given for each model.

<table>
<thead>
<tr>
<th>Intercept</th>
<th>Distance mainland</th>
<th>Age</th>
<th>Age²</th>
<th>Area</th>
<th>Nearest island</th>
<th>k</th>
<th>AICc</th>
<th>ΔAICc</th>
<th>AICc weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>−0.3117</td>
<td>0.2410</td>
<td>−0.0129</td>
<td>0.4466</td>
<td></td>
<td></td>
<td>9</td>
<td>282.6</td>
<td>0.00</td>
<td>0.51</td>
</tr>
<tr>
<td>−0.4606</td>
<td>0.2336</td>
<td>−0.0129</td>
<td>0.4639</td>
<td>0.0018</td>
<td>10</td>
<td>284.3</td>
<td>1.7</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>−0.7650</td>
<td>0.0003</td>
<td>0.2518</td>
<td>−0.0132</td>
<td>0.4535</td>
<td>10</td>
<td>284.6</td>
<td>2.0</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>−0.8565</td>
<td>0.0002</td>
<td>0.2433</td>
<td>−0.0131</td>
<td>0.4693</td>
<td>0.0017</td>
<td>11</td>
<td>286.4</td>
<td>3.8</td>
<td>0.08</td>
</tr>
</tbody>
</table>

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 ATT² 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 archipelago 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.

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 (km²) and age of the island (n = 134). The grid is predicted from the fixed effects estimates from the model with the lowest AICc (Table 4) adjusted by 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.
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 (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.

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 a mean of zero and 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 archipelagos.

Acknowledgements – 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 and R. J. Whittaker 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


Glossary

Crossed random effect – two or more random effects that act independently, such as island and taxa.

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.

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 (ΔAICc) and AICc weight is given for each model. The random effects for the models were island, taxon and archipelago. Archipelago (ARCH) was included as a fixed effect and its interactions with age, age² 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.

<table>
<thead>
<tr>
<th>ARCH</th>
<th>Age</th>
<th>Age²</th>
<th>Area</th>
<th>Nearest island</th>
<th>ARCH: age</th>
<th>ARCH: age²</th>
<th>ARCH: area</th>
<th>k</th>
<th>AICc</th>
<th>ΔAICc</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>14</td>
<td>271.5</td>
<td>0.0</td>
<td>0.24</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>11</td>
<td>272.2</td>
<td>0.25</td>
<td>0.17</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>17</td>
<td>272.9</td>
<td>1.43</td>
<td>0.12</td>
</tr>
</tbody>
</table>


