

Homework 12

This assignment will use some data from the attached paper by Cabral et al. The authors looked at biogeographic and environmental variables that predict the species richness of vascular plants on oceanic islands. Asking what predicts species richness on oceanic islands is a classic question; the more novel angle of the authors was to focus on the fact that islands often come in archipelagos, and there are likely important processes that operate at the archipelago scale, in addition to the island scale. We'll use this idea as an example for using mixed models.

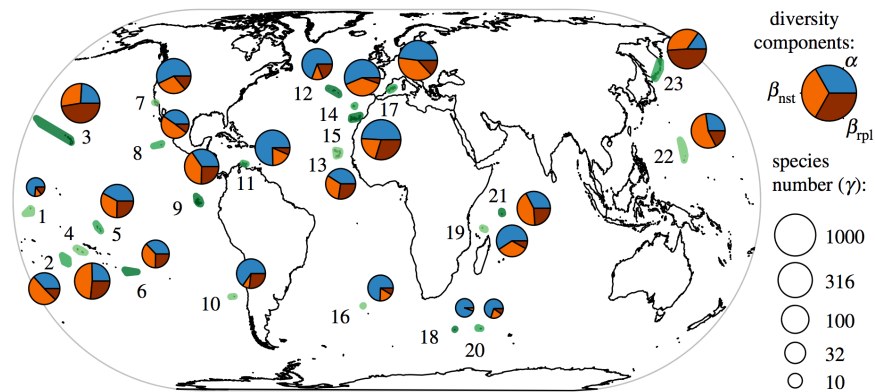


Figure 1. Archipelagic plant richness (γ) and α , nestedness (β_{nst}) and replacement (β_{rpl}) components. The vascular plant floras of the 23 archipelagos analysed are highlighted in green: 1, Phoenix Islands; 2, Cook Islands; 3, Hawai'i; 4, Society Islands; 5, Marquesas; 6, Pitcairn Islands; 7, Northern Californian Channel Islands; 8, Revillagigedo Islands; 9, Galapagos Islands; 10, Juan Fernandez Islands; 11, Dutch Caribbean; 12, Azores; 13, Cape Verde; 14, Madeira; 15, Canary Islands; 16, Tristan da Cunha; 17, Balearic Islands; 18, Prince Edward Islands; 19, Aldabra; 20, Crozet Islands; 21, Inner Seychelles; 22, Marianas; 23, Kuriles.

The authors compiled a database of many variables, which I've simplified for this exercise. The file 'cabral_island_data.csv' contains these variables measured at the island scale: species richness ('Species'), island area (km²), maximum elevation (meters), and mean temperature. The file 'cabral_arch_data.csv' contains these variables measured at the archipelago scale: age of the oldest island (millions of years), number of islands in the archipelago, distance of the archipelago from the nearest continent (km).

All of these predictors could be important for species diversity on islands, and the point of the analysis is to figure out which ones actually have support, using a comparative approach. We will focus on island-level richness (the authors also look at between-island beta diversity and whole-archipelago gamma diversity). Island area is a proxy for environmental heterogeneity, and also affects the target size for colonists; maximum elevation is also a proxy for heterogeneity; mean temperature may or may not be important, depending on how you think the latitudinal diversity gradient arises; age of the oldest island determines time for colonization and time for diversification; number of islands in the archipelago could determine the total target size for colonization; and distance to the mainland will affect the supply of propagules.

To start, let's consider that the data are naturally organized into groups (archipelagoes), and we should account for this structure (non-independence) in the model. In addition, some of the predictors are defined at the scale of the archipelago, which means we need an archipelago random effect in order to not pseudoreplicate when testing those predictors. Make a model with species richness as the response, and with a random effect for Archipelago. Species richness could be modeled as discrete count data (e.g., a negative binomial distribution), but we'll cover GLMMs later; for now you can use $\log(\text{Richness}+1)$ to get a pretty normal looking response. What proportion of the variation in species richness occurs at the archipelago scale, and what proportion occurs within archipelagoes? Which archipelagoes are particularly diverse, and which are depauperate?

Now let's think about the six predictors. Make some exploratory plots of the effect of each variable on richness, plotted at the appropriate scale. You'll need to merge the datasets. Think about which predictors might need to be transformed for use in a linear model.

Make a mixed model that includes all the predictors. Plot fitted effects (fixed and random), plus model diagnostics. How does it look? How much variation do the predictors explain, at the two different scales in the data (island and archipelago)? Present R^2 values for the two scales. Also, how much of the total variation have they explained, according to $R^2_{\text{GLMM}(m)}$?

Use approximate F-tests as described in lecture to test the significance of the predictors. How do you interpret the results of these tests, and the effects plots, in light of hypotheses for what controls species richness in islands and archipelagos? What are the denominator degrees of freedom for each predictor? This is essentially telling you how much replication there is for that predictor, minus the number of parameters for that predictor. Do the denominator df make sense? Why or why not?

Is the model we've used the best model? Often I just stick with one big model when the ratio of data to parameters is pretty good. But some people really like to find the best model, or to look at model weights, etc. Use AICc in some capacity to assess which predictors are important, what the 'best' model is, and how sure you are about what the best model is. The details of how you do it are up to you, as long as it seems justifiable. Remember to do `REML=FALSE` for comparing models.