MA49 Community ecology in Madagascar

In Mahamavo dry forest in western Madagascar we monitor patterns and trends in biodiversity of flowering plants and terrestrial vertebrates using standard methods. It would be possible to join one of the field science teams, contribute to data collection, and then take data from our database to undertake a community analysis.

Should you decide to do a dissertation on community ecology, I would recommend focusing on just one of the following taxonomic groups:
- flowering plants
- birds
- reptiles and amphibians
- mammals

However, before beginning to think about analysis of ecological communities in Mahamavo, it's important to understand some key aspects of the field methods, particularly our sampling scheme.

Sample unit = place in the forest of known co-ordinates at which a scientific sampling activity takes place eg the named approx 3km long sample routes used for monitoring lemurs and herps. Also, this applies to point locations used for bird point counts.

Occasion = a particular time on which a scientific sampling activity takes place. Typically, each sampling unit will be sampled on many occasions in a field season.

Record = an observation made on a specified occasion. On each occasion, usually several records will be made, but it's possible that there might be no records on an occasion.

This creates a nested data hierarchy which can be manipulated in various ways. The most obvious things we do are to bind all the sample units with the occasions and then bind all the records with the occasions to create a very rich table which tells you the full context of every observation. It's also possible to work out things like detection histories, convert records into effort-corrected relative abundances, etc.

The potential problem for community analysis of animals surveyed on routes (herps, mammals), although not on points (birds) is that the number of sampling units (routes) that will be available for you to analyse will be quite small (around 20). Instead you may wish to split these units up spatially by resampling the database into around 150 grid squares to create more unique sample units each with a robustly observed animal community.   This is possible because every observation is spatially referenced by taking co-ordinates with a GPS receiver. The reason for disaggregating data like this is to make later statistical tests more powerful. As you can see, the whole route has a community of 10 blue, 2 green and 3 red animals but each grid square also has a distinct animal community eg 2 blue and 1 green.

300 m or 1km
squares

animal observations
colour coded by
species

Sample route
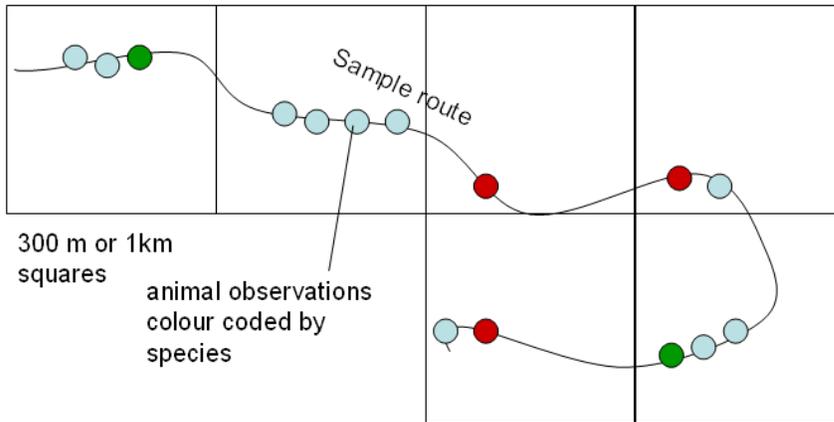
## Possible analysis methods

Here's a list of potential approaches to think about and guide your reading. A good project would probably explore the data using 2-3 of these methods.   These are in rough order of difficulty, from childishly simple to mind-bendingly mathematical:

- Plot a rank abundance curve and interpret it
- Rarefaction analysis to estimate species richness and evaluate efficiency of sampling
- Calculate community attributes at the level of each sample unit and model them in relation to the environment
- Co-occurrence tests for ecological interactions
- Pairwise analysis of beta-diversity
- Use species distribution models to look at community structure
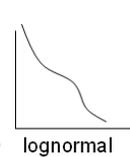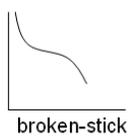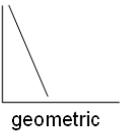- Generalised Dissimilarity Models

Further details of each…

Rank-abundance curves

A rank abundance curve generally looks like this…
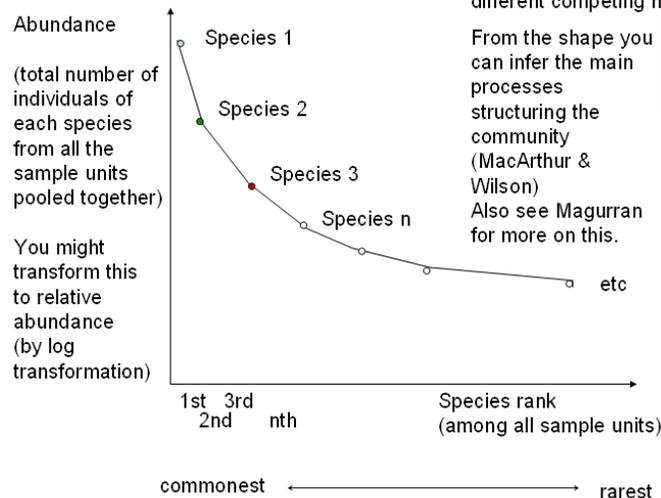
The shape is very informative about the community. You can test which it is for a particular community by fitting different competing models.

From the shape you can infer the main processes structuring the community (MacArthur & Wilson)
Also see Magurran for more on this.

Abundance

(total number of individuals of each species from all the sample units pooled together)

You might transform this to relative abundance (by log transformation)

Species 1

Species 2

Species 3

Species n

etc

geometric

broken-stick

lognormal

etc

1st   3rd
   2nd   nth
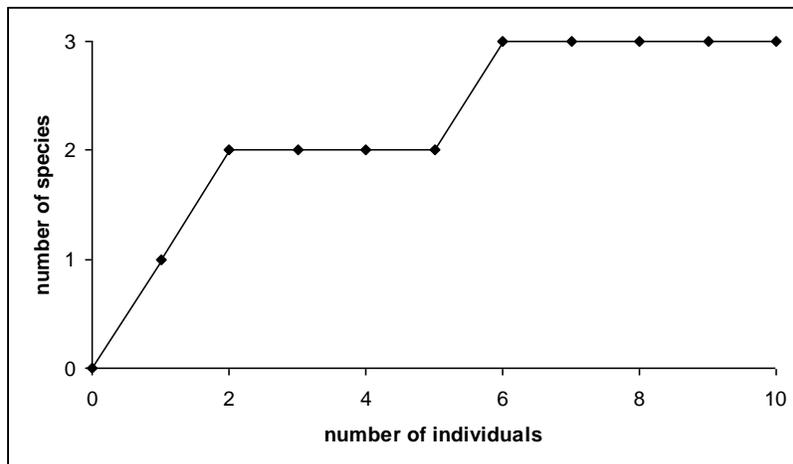
Species rank
(among all sample units)

commonest ⟵⟶ rarest

Rarefaction curves

For each particular taxonomic group such as birds which we have been sampling in Mahamavo, we have a species list and can say that our recorded species richness is say 110 different bird species. How do we know whether this is the real species richness of the forest or whether if we keep looking harder we might record further different species in the site? Basically you need to plot and interpret rarefaction curves. The procedure for this is very simple. We use our database to list all the records of individuals for a given taxonomic group (birds, reptiles, mammals, flowering plants) in date order and then plot the cumulative frequency of newly observed species.

Like this:

Record number	Species	Cumulative number of species
1	A	1
2	B	2
3	B	2
4	A	2
5	A	2
6	C	3
7	A	3
8	B	3
9	A	3
10	A	3
and so on…		

Then plot, like this

The fact that the curve seems to have reached a plateau suggests that the richness in this (example) community is just 3 species and that no matter how much harder we look, we won't find any more. However to be sure, we must use k-fold partitioning or bootstrapping on a much larger data set to

partition the data into lots of random sets and plot the lots of curves with an error distribution in order to be sure whether the gradient is flat or not. Additionally, if a curve has not yet reached an asymptote, it is possible to estimate, with an error, what the true richness actually is. The curve on the left is a group which probably has reached an asymptote. The curve on the right represents a group in which new species for the site are still being found.
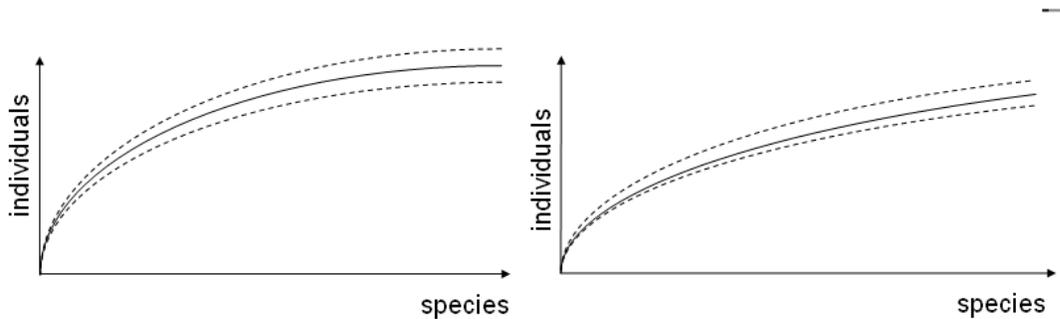
It would be particularly interesting to use the full database spanning 2009 to 2014 to plot rarefaction curves for several different groups eg plants, birds, mammals, reptiles. I would expect that the species list is nearly complete for birds and mammals, but still very incomplete for herps and plants.

Community attributes

Community attributes are emergent properties of the community that you can calculate from all the observations on a particular sample unit. The simplest is species richness, but also diversity metrics such as Shannon-Wiener, Simpson, etc. However, the subtlety is that as you expend more effort collecting data on a sample unit on more occasions, the estimated community attributes will change. There will also be area effects depending on the size of the units. Eg species richness N will slowly drift up as species accumulate (you tend to see the common species immediately, and only detect the rarer ones after a bit of effort). Same point applies to estimates of diversity metrics.
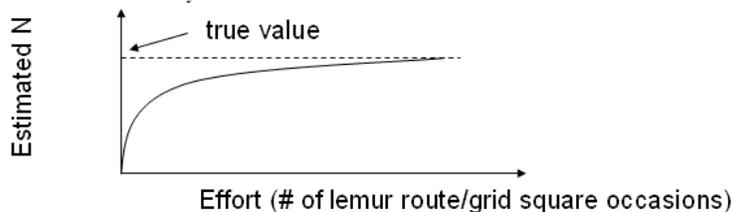
How to deal with this?
1. standardise. Expend the same amount of effort on all sample units. OK, but drawback is that the units vary and the optimal effort may differ, also hard to achieve.
2. Rarefaction analysis. Use bootstrapping to model the shape of the curve, find N
3. Estimate true occupancy status of each sampling unit by each species from detection histories and site covariates, accounting for variable prevalence and detectability.

What can you do with community attributes?

Imagine you have queried the database and GIS to make a data table like this…

Maybe corrected in some sophisticated way

Unit	Sp rich	Diversity	Site-specific covariates (from GIS)			
			eg elevation	water	climate	etc
1						
2						
3						
…						
n						

Now frame a set of competing hypotheses to explain the pattern of variation in community attributes across the set of sample sites eg…
- Species richness is higher at lower elevations
- Species richness is greater in larger sized forest patches

Then use simple statistics eg. regression or multiple regression to decide which is the most parsimonious explanation from the set of competing hypotheses.

Co-occurrence tests

Very simple to do, but note this is quite an old-fashioned method. A more modern approach is generalised dissimilarity models.

Here's the general approach:

Cross-tabulate presence or absence of every species on every sample unit.

Site	Species			
	1	2	3	n
1	0/1			
2				
3				
n				
Σ				

Now, for every possible pair of species count the frequency of sites in the four possible contingencies of pairwise-combinations of both being either present or absent at a site:

- Do a Fisher's exact test on this table.
- Repeat for all pairs of species ie do $n^2 /2 - n$ tests.
- A significant result implies an ecological interaction eg competition.
- Need to be very careful with significance thresholds as many tests are being done.
- Tabulate the results of the tests into a triangular matrix.
- Look for patterns. Do any species have lots of significant interactions? These are potentially keystone species.
- Construct a network of putative ecological interactions. Edge weights derived from the test statistics.
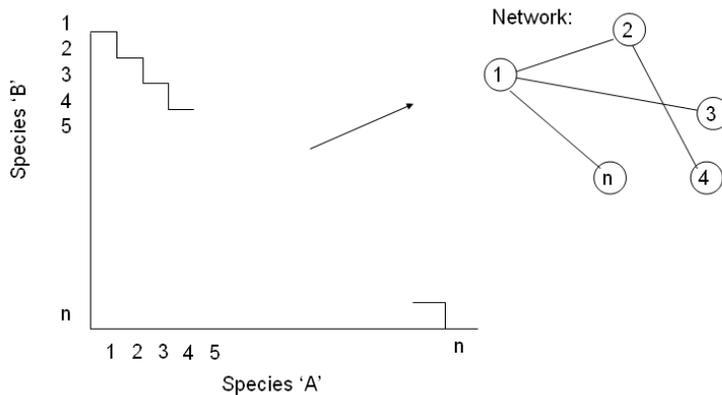

Pairwise analysis of beta diversity

The sample units in this analysis would be the sample units (routes or grid squares). First, tabulate presence-absence for each species for each sample unit or tabulate effort-corrected relative abundance (RA) for each species for each sample unit to make a table like this.

Site	Species			
	1	2	3	n
1	RA			
2				
3				
n				

Now calculate a pairwise community similarity metric eg Jaccard (if pres-abs) or Bray-Curtis or Chao statistic (if RA) for every possible pair of sites. This is most easily accomplished with the program Estimates http://viceroy.eeb.uconn.edu/estimates

Note that the index that you can calculate will depend on whether you know presence/absence of species on sites or whether you have a count or relative abundance of individuals of each species on each site. Jacquard uses presence/absence whereas indices such as Bray-Curtis and Chao use counts. This is also relevant if you decide to use the outputs form a set of hardened distribution models to assess community similarity in which case you would be limited to using just the Jacquard index.
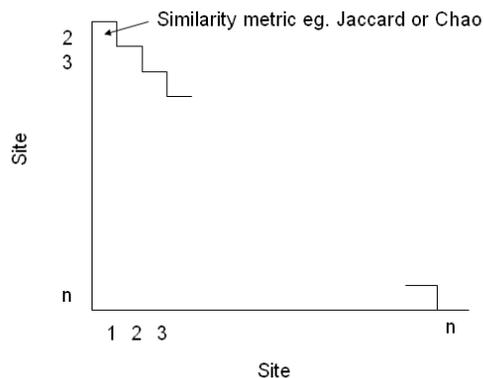

Now, set up a series of 4 or 5 competing hypotheses for why pairs of animal communities might be similar and then use your data and analysis to identify which hypothesis or hypotheses are most likely to be correct

Eg.
- Physically close communities will be more similar (spatial autocorrelation)
- Communities in more physically similar habitats will be similar.
- Area effects (island biogeography)
- Presence or absence of particular keystone/foundation species eg Fig trees affect the community composition.
- Level of human disturbance determines which species occur in a community

Collect data about the sites (from GIS) and use this to calculate pairwise matrixes of similarity between every pair of sites in terms of covariates relevant to your hypotheses. Eg make a pairwise Euclidean distance matrix.
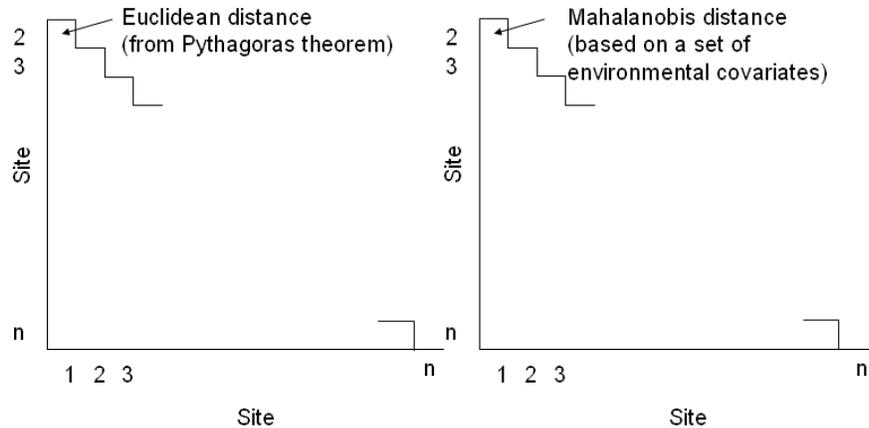
Now, test your competing hypotheses by using Mantel tests between the community similarity matrix and the other matrixes representing the explanatory variables. Mantel tests are most conveniently done in R using the package VEGAN (for vegetation analysis – not veganism). You should probably should also use an information-theoretic approach to your model selection (Burnham and Anderson 2004).

You might think that you could just do a linear (or multiple linear) regression between all the elements in your community dissimilarity matrix and the complementary elements in the other matrices. However this would be very very wrong. The reason is that if you have n sample units there will be $n^2/2-n$ elements in the triangular matrix – many more elements than n. If you did regression on this you would massively inflate the estimated number of degrees of freedom, and falsely get significant results. Mantel (and partial Mantel) tests overcome this problem by permuting the response matrix and doing a regression on just n randomly selected elements and their complementary elements from the other matrixes, then repeating the permutation thousands of times to test for significance properly with the right number of degrees of freedom.

Use species distribution models to look at community structure

Join spatial presence records for each animal species with environmental data (from GIS). Use GLM or Maxent to make and then validate distribution models for your chosen species in Mahamavo. See the guidelines on distribution models for more information on how to do this.

Now for the community ecology. Analyse the hardened habitat suitability maps. I can think of several interesting approaches (ordered from very easy to quite hard):
- Add them together to make lovely colourful map of species richness in a taxonomic group.
- Numerically classify (just like with RS data) to make a map of community types.
- Do ordination on this stack of species maps to make a 3-colour composite map of the first 3 axes of compositional variation.
- Derive a map of beta diversity by passing a custom moving window over the stack of grids (eg McKnight 2007).

- Calculate the lag of the semivariogram for each species, resample each map by thinning at the lag distance to remove this effect. Now use this very large dataset with environmental covariates to do a more powerful pairwise analyses in exactly the same way as described earlier for the sample sites (except you would have to use the Jacquard index as the response).

<u>Generalised dissimilarity models</u>

Generalised dissimilarity models require that you aggregate the biodiversity records to a number of sites of known x y co-ordinates. These could be bird point count sample sites or a grid of squares imposed onto the sample routes or even all unique spatial records (including opportunistic records) for a group. You then tabulate presence or absence of each species on each unit and use the GIS to find the values of as many environmental covariates as you want to consider eg elevation, vegetation properties, disturbance, climate (see the SDM guide for much more information on covariates).

unit	x	y	sp 1	sp 2	sp 3	sp n	cov 1	cov 2	cov 3	cov n
1			0/1							
2										
3										
n										

You also need maps of all the covariates across the extent of your chosen study area (already made).
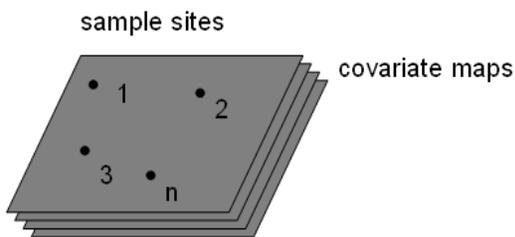
GDM is a generalised (ie non-linear) matrix regression technique which relates pairwise dissimilarities between a limited set of sample units with environmental dissimilarities between the same sites measured using a set of covariates with continuous spatial coverage across the study area (this is the important distinction between the pairwise analysis with Mantel tests and GDM). The other major advantage of a generalised approach to dissimilarity modelling is that it becomes possible to account for non-linear relationships between environmental features and dissimilarity and also to account for the fact that most indices, like Bray-Curtis are constrained to a 0-1 range, so there is the possibility of asymptotic relationships. (Ferrier et al 2002; 2006)

Accomplishing this analysis is only possible in R, but I have made scripts to do it painlessly. It is done in two parts. First a dissimilarity object is constructed, this is the easy bit. Second you back-project this object in onto the maps of environmental covariates to make a map of modelled beta diversity across

the study area. The tricky bit here is that it is computationally intensive so you must loop this code over many small tiles. These maps look sophisticated and colourful and also have utility for conservation planning in the sense that land with compositional turnover or high beta-diversity, which has hitherto been hard to map, might be high conservation priority. It might also be interesting to interpret such maps in relation to other landscape features.

Another interesting dimension to this would be to compare the pattern of beta diversity generated from field data with the pattern generated from GBIF records from the Mahamavo region.

Suggested Reading

Burnham KP, Anderson DR (2004) Multi-model inference: an information-theoretic approach. 2nd edition. Springer

Ferrier SM, Drielsma, et al. (2002). Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. II. Community-level modelling. Biodiversity and Conservation 11(12): 2309-2338.

Ferrier SM, Guisan A, (2006) Spatial modelling of biodiversity at the community level. Journal of Applied Ecology 43, 393-404.

McCune, Grace (2000) Analysis of Ecological communities. MJM

Verhoef HA, Morin PJ (2010) Community ecology. Oxford

R VEGAN package handbook
http://cc.oulu.fi/~jarioksa/softhelp/vegan.html

Colwell – Estimates website
http://viceroy.eeb.uconn.edu/estimates