MADAGASCAR DISSERTATION/THESIS PROJECT

Modelling lemur distributions in Madagascar

Species distribution modeling is a technique for linking spatially referenced records of species occurrence – for example collected during appropriately designed field-based biodiversity monitoring programmes – with maps of environmental variables such as elevation, climate, vegetation or human disturbance, in order to create a statistical model of the relationship between a species and its environment, ie the species realised ecological niche. GIS can then be used to express the results of models as habitat suitability maps across a desired spatial extent. These output habitat suitability maps are also a powerful tool to communicate conservation messages to non-scientists

The species records may be a set of presences only or a set of presence and absence records, depending on the detectability of the species and sampling method that has been used. Almost any environmental variables can be used in a distribution model, although it is normal to select a restricted set of variables at a particular spatial scale based on a working hypothesis about the aspects of the environment which may be important to the focal species.

A wide range of statistical approaches have been developed for fitting distribution models including various types of regression models, machine learning and classification methods. Regression approaches such as GLM are simple to implement; however, more complex information theoretic approaches, especially maximum entropy, have proved to be very powerful. Once a model has been built and refined, it is critically important to validate the model on a subset of data which was not used in model construction, in order to objectively assess how well the model performs. Please see the slides at the end of this document for an overview of the modeling process.

Species distribution models is a popular dissertation topic at the Madagascar site. There is scope for great flexibility in the ways that this technique can be used and there are many possible extensions and additional analyses that are possible. Here, we will focus on lemurs.

The aim of repeatedly walking lemur routes in the day and at night is to estimate encounter rates, densities and population sizes of diurnal lemurs using distance sampling. The commonest diurnal lemurs is Coquerel's sifaka (*Propithecus coquereli*), but you will also see common brown lemur (*Eulemur fulvus*), mongoose lemur (*Eulemur mongoz*), At night there are hundreds of mouse lemurs; golden-brown mouse lemur (*Microcebus ravelobensis*) and grey mouse lemur (*Microcebus murinus*). You might also see Western Avahi (*Avahi occidentalis*), Milne-Edwards sportive lemur (*Lepilemur edwardsi*), Western fork-marked lemur (*Phaner pallescens*), and Fat-tailed dwarf lemur (*Cheirogaleus medius*).

Extensions

Additionally, you need to decide whether to undertake a straightforward analysis of a set of species or whether you want to use the results for further analysis, in which case you would need to decide where to place the emphasis in your analysis. You may wish to use distribution models to look at species responses to landscape configuration, or use the results to explore patterns of biodiversity and even hence undertake systematic conservation planning. Maybe you might decide to investigate covariate spatial scale effects by comparing models made with, say 30m and 300m resolution data. A related approach is to create focal covariates to explore the scale at which each species responded to landscape features. You might consider looking at the effects of historical environmental changes or future change (eg climate change, land cover change) on species distributions. Perhaps you would be interested to contrast the performance of several different modeling techniques such as Maxent, GLM, ENFA. Alternatively, you might want to look at the feasibility of assimilating new data (eg RS) to generate indicators e.g. of area of occupancy, for use as an aid to monitoring. It could also be possible to generate detection histories for sample units, model detectability and correct distribution models for imperfect detections. A final, very neat possibility is to use records for Mahamavo from GBIF to make distribution models and compare estimated attributes such as patterns of species richness estimated from field data with that from web databases or to evaluate the accuracy of polygons denoting the global ranges of endangered species and investigate statistical range polygon refinement methods.

These more advanced topics may sound quite complicated, so please don't hesitate to get in touch to discuss further and ask questions

Practical aspects: fieldwork, model type, covariates

In practical terms, I would recommend that when you come to Mahamavo, you join in with one of the biodiversity survey teams and help with general data collection for a few weeks, then query our field database (covering five years from 2009 to 2014) to generate your dataset, which you then join with whichever spatial data you want to use. I maintain all the spatial datasets for Mahamavo, so you will not need to prepare these (covariate preparation is normally the hardest part of distribution modelling), although if there's extra covariates that you want to incorporate, I can show you how to do this.

Full training will be given in the necessary data handling and the complete modeling procedures on site, including validation statistics, cartography etc.

Whichever taxonomic group you choose and regardless of whether you want to pursue one of the many possible extensions, there will be lots of choices to be made about the modelling approach to take in the first place. I recommend using either GLM,a regression method, and/or Maxent, an information-theoretic method. GLM is conceptually simplest, fairly straightforward to do and usually gives very good results. The mathematical basis of Maxent is harder to understand, but in practical terms, it's very easy to make models and they generally perform slightly better than GLM.

You will need to decide on the spatial extent and the scale (spatial grain) of your modelling. (eg the whole Mahamavo watershed at 30m resolution), and also think about the a set of environmental covariates to use which will be relevant to your taxonomic group.

Available covariates at both 30m and 300m scale for Mahamavo include:
Reflectance derived (from Landsat 5 and & at 30m, from MERIS and MODIS at 300m)
- NDVI –normalised difference vegetation index
- EVI – enhanced vegetation index
- Tasseled cap greenness, moistness, brightness – indices of healthy green vegetation
- BRDF volumetric and geometric scattering kernel parameters – proxies for canopy architecture
- Texture metrics – proxies for habitat heterogeneity

Elevation derived (from SRTM)
- Elevation
- Slope
- Aspect and sin(aspect) and cos(aspect) as aspect is a circular variable
- TWI - Topographic wetness index
- Relative insolation

Distance metrics (from 1:100k topo maps)
- Distance to water
- Distance to road
- Distance to village

Configuration metrics (from a classification of
- Patch size
- Patch perimeter:area ratio
- Distance from patch edge
- Patch isolation

Soils (from 1:100k soils maps)
- Soil class

Climate (from WORLDCLIM station interpolation and also RS observed from TRMM, DMSP-SSMI, AIRS etc)Several climate parameters are available eg mean annual temp, total annual precipitation but please note that the study area is relatively small, with limited elevation range. As such most climate parameters show very limited variation across the study landscape.

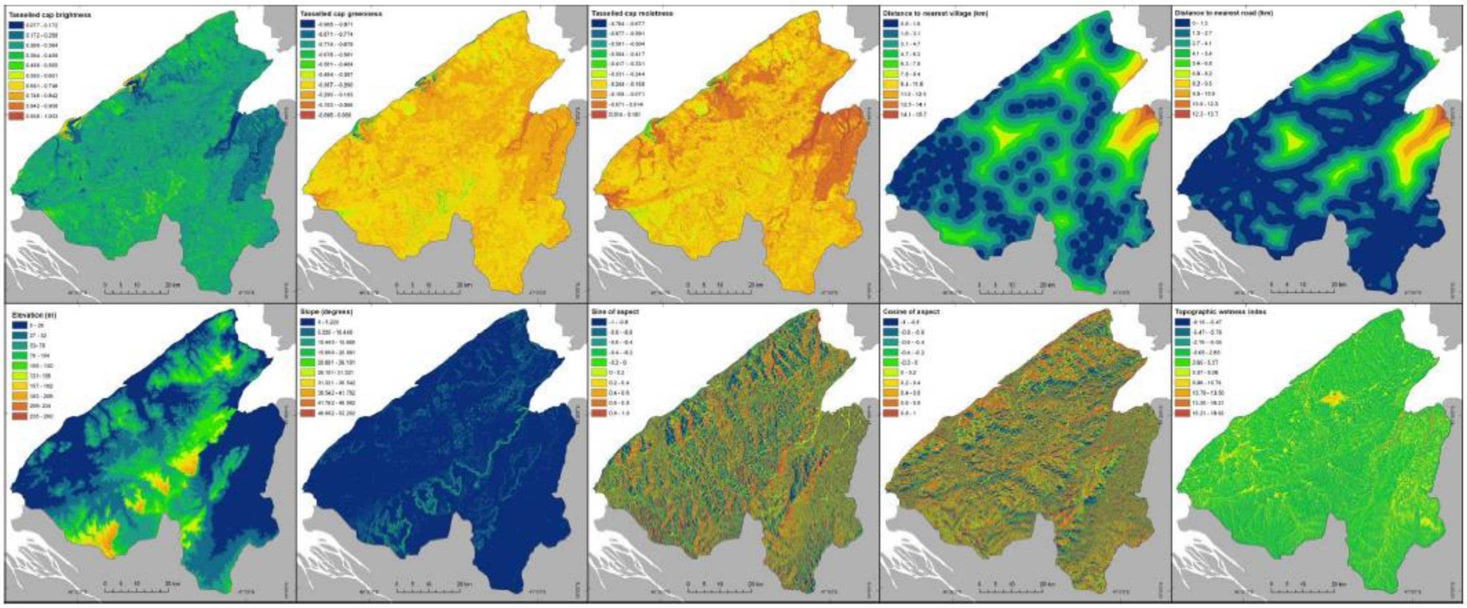

Figure. Maps of some covariates for Mahamavo at 30m resolution: TC1,2,3, village distance, road distance, elevation, slope cos(aspect), sin(aspect), TWI.

Some of these variables are static, such as elevation, whereas others like reflectance and configuration change with phenology and land cover change. It is possible to create covariates to capture these effects, such as the first and second principal components of a monthly time series of NDVI, EVI or TC1 to capture mean greenness and seasonality in greenness.

At this stage, please remember that the final way you do your project will be shaped by your reading and your interests and you don't actually need to decide the answers to all of these issues immediately.

Reading:

Fielding, Bell (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24 (1): 38–49

Guisan, Thullier (2005) Predicting species distribution: offering more than simple habitat models. Ecology Letters *8*: 993–1009

Gusian, Zimmerman (2004) Predictive habitat distribution methods in ecology. Ecological Modeling 135: 147-186

Hirzel, Guisan (2002) Which is the optimal sampling strategy for habitat suitability modeling? Ecological Modeling 137: 331-341

Liu et al (2005) Selecting thresholds of occurrence in the prediction of species distributions. Ecography 28: 385-393

Pearson (2004) Modeling species distributions in Britain: a hierarchical integration of climate and landcover data. Ecography 27: 285-298

Franklin (2009) Mapping species distributions. Cambridge
Phillips (2008) Maxent handbook
Moilanen, Wilson, Possingham (2009) Spatial conservation prioritisation. Oxford
Margules & Sarkar (2007) Systematic conservation planning. Cambridge
Moilanen & Kujala (2008) Zonation manual