Identifying suitable habitat, uncertainties, and pre-import decisions

Session: *Science and Economics of Pre-Import Screening for Animal Species*

Simon C Barry[1]

Invasive species represent a significant threat to the world economy and environment. Because of this threat there is interest in exploring ways of predicting when a particular species will be invasive in a new environment. This information can then be used by policy makers as they consider the relevant issues in their jurisdiction.

Reviews of studies to predict invasiveness have been carried out by Kolar and Lodge (2001) and updated by Hayes and Barry (2008). A significant body of literature has built up across a range of taxa looking at correlates of invasion success. Several points can be made from this literature. First, while invasiveness cannot be predicted with certainty, statistical approaches can produce models that on average provide significant discrimination, and characterise the nature of this uncertainty. Second, while a range of factors are related to invasiveness in particular circumstances, few consistent predictors have been found. In summary climatic match, history of invasiveness and number of releases/size of release are consistently related to establishment success.

This result is not particularly surprising biologically. The factors that are important to success will critically depend on the context. The impact of climate on species is not controversial, and the effect of release size and number of releases is clearly interpretable. In considering the risk of an importation, the habitat suitability is therefore an important component of assessing the likely impacts of the species if it escapes captivity**.**

Approaches to predicting habitat suitability have developed significantly over the last 30 years. The original approaches such as BIOCLIM (Busby, 1991) and DOMAIN (Carpenter *et al*, 1992) sought to define regions in the environmental space defined by measured variables where the species was present. While significant in their time, these approaches suffered from dimensionality problems and could not adequately deal with complex patterns in environmental space.

Advances came by viewing the problem as a regression problem: what was the probability that a species is present given the measured environmental conditions. This developed from the use of simple generalised linear models to generalised additive models and regression tree approaches in the 1990's (Austin and Meyers, 1996). These models were an improvement because they were more flexible and data driven that the simple GLM's.

In the previous seven years there has been an acceleration in the use of techniques from the statistics and machine learning communities and improvements have been significant. Boosted regression trees and Maximum Entropy modelling (Phillips et al, 2006) have been particularly successful. The modern techniques are better at modelling complicated interactions in the data. They do not impose as rigid

[1] CSIRO Mathematics and Information Sciences, simon.barry@csiro.au.

constraints and assumptions such as linearity and additivity and provide a better trade off between flexibility and over-fitting.

This discussion is reflected in the results of a landmark empirical study on the performance of species prediction models. Elith *et al* (2006), in a major collaboration between practitioners in the field, assessed the empirical performance of a range of approaches. They demonstrated that on average modern machine learning techniques out performed other approaches.

This presents an attractive path for pre-import screening. The distribution of a species could be modelled in its current range and this model could then be projected onto the potential recipient region and the results could be used in the decision making. Issues with this approach arise.

First, while techniques have improved there are still errors and uncertainties in almost all predictions made from models. Barry and Elith (2006) have considered these. They typically arise from the effects of significant variables not available for the analysis and from global misspecifications in response structure. These errors and uncertainties can be intractable. They represent failure in the amount of available information and cannot be rectified by a model either in the systematic or random component. They are typically not spatially or environmentally homogeneous.

In addition there are a number of particular issues that arise:

Novel climates: Particular climatic conditions that are conducive to the persistence of the species may not occur in the established range and so there is no information to test the suitability of these climates.

Biotic interactions: Species distribution is a combination of the theoretical biological niche, historical effects and other biotic interactions. Thus the established range may overestimate/underestimate the potential range in a new site when other biotic interactions are present.

Taxonomic problems: The question about which records of a species are relevant to a particular introduction attempt will arise. Depending on the effort of taxonomists species can have large or small ranges. How this relates to translocation success is understood in only a limited number of examples.

Data quality: For many species the available presence records may be limited. Data can refer to entire countries or have suspect elements. The most time consuming and difficult part of the analysis is the generation of quality assurance of the data. Thus global attempts to standardise and collate such data are important.

Summary statistics: After the analysis the question of how to best summarise a prediction and how to bring this into the decision making process is important.

While the previous discussion may highlight the issues in the use of the techniques it should not be taken as an argument against them. For all of these limitations there is significant evidence that quite coarse approaches to climate matching will improve the ability to predict invasiveness (for example Forstyth *et al*, 2006). What should be avoided are narrow arguments that one technique is better than another, undermining confidence in the system. Effort is better expended developing constructive approaches to habitat suitability modelling. These would set out clearly the steps in the analysis so that different individuals with the same information would come to approximately the same conclusion. This would cover treatment and assessment of data, validation and modelling technique and provide a key building block for a transparent, science based decision making system.

References

Austin, M.P. and Cunningham, R.B. 1981. Observational analysis of environmental gradients. *Proc. Ecol. Soc. Aust.* 11: 109_ 119.

Austin, M.P. and Meyers, J. 1996. Current approaches to modelling the environmental niche of eucalypts: implication for management of forest biodiversity. *Forest Ecol. Manage.* 85 pp. 95–106.

Barry, S.C. & Elith, J. 2006 Error and uncertainty in habitat models. *Journal of Applied Ecology* 43,413-423

Busby, J.R. 1991. BIOCLIM - a bioclimate analysis and prediction system. _ In: Margules, C. R. and Austin, M. P. (eds), *Nature conservation: cost effective biological surveys and data analysis.* CSIRO, pp. 64-68.

Carpenter, G., Gillison, A. N. and Winter, J. 1993. DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodiv. Conserv.* 2: 667-680.

Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R. J., Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. McC., Peterson, A. T., Phillips, S. J., Richardson, K. S., Scachetti-Pereira, R., Schapire, R. E., Soberón, J., Williams, S., Wisz, M. S. and Zimmermann, N. E. 2006. Novel methods improve prediction of species' distributions from occurrence data. – *Ecography* 29: 129–151.

Forsyth D.M., Duncan R.P., Bomford, M., Moore, G. 2004. Climatic suitability, life-history traits, introduction effort, and the establishment and spread of introduced mammals in Australia. *Conserv Biol* 18(2):557–569

Hayes, K.R. and Barry, S.C. 2008. Are there any consistent predictors of invasion success? *Biological Invasions* 10:483-506

Kolar C.S., Lodge D.M. 2001. Progress in invasion biology: predicting invaders. *Trends Ecol Evol* 16(4):199–204

Phillips, S. J., Anderson, R. P. and Schapire, R. E. 2006. Maximum entropy modeling of species geographic distributions. Ecol. Modelling 190: 231_ 259.