

OLM 5.2. Correlative distribution models

Characterisation of the geographical distribution of species by climatic, geomorphological and biotic data provide opportunity to deduce ecological tolerance ranges for species (Ch5.2.2, p.84) and to predict expected changes in these distributions (Ch5.2.1, p.77). Besides the advantage of having huge on-line databases that can support this type of analysis there is quite a number of pitfalls that have to be taken into consideration at the interpretation of the results. We discuss some of them in this OLM.

Correlative methods are widely used for exploring species tolerance ranges. These methods are called species distribution modelling (Elith and Leathwick 2009) or niche modelling (Peterson 2006), the latter referring to the Grinnellian niche concept (Ch5.2, p.86; Ch10.2.2, p.209; OLM 10.4). Thanks to the bulk of data obtained from museum collections, species distribution maps and extensive standardized climatic and geomorphological measurements made on a regular basis, the correlative method is capable of providing information on the abundance-based range of tolerance (Ch5.2.2, p.84) to many different environmental factors for many species,. However, the quality and the form of these data types differ. Distribution data from museum collections do not reveal absences, which is problematic because only presence-absence data can be used for estimating the probability of occurrence. Therefore analysis of presence-only data requires special methods, or they are often complemented by assumed absences (so-called *pseudo-absences*) before applying methods suitable for presence-absence data. Maps of species distributions – when based on systematic surveys – provide presence-absence data. Environmental (climatic, soil, geomorphological, land use etc.) data are available from dedicated GIS databases into which species occurrence data are also fed (Figure 5.2.1). The comparison of model predictions with field data not used in model building serves as feedback for testing and improving the model. Correlative distribution models can be used to forecast the occurrence of species at locations that are either poorly surveyed locations or under climate change, and they may even facilitate the discovery of related but unknown species, with similar tolerances to that of the focal species. Seven endemic species of chameleon have been discovered on Madagascar in this way (Raxworthy et al. 2003). Correlative models can be applied to forecast the effects of climatic shift (e.g. Czúcz, Gálhidy and Mátyás 2011, Thuiller et al. 2011) as well as the spread of invasive species (e.g. Aragón, Baselga and Lobo 2010, Dullinger et al. 2009).

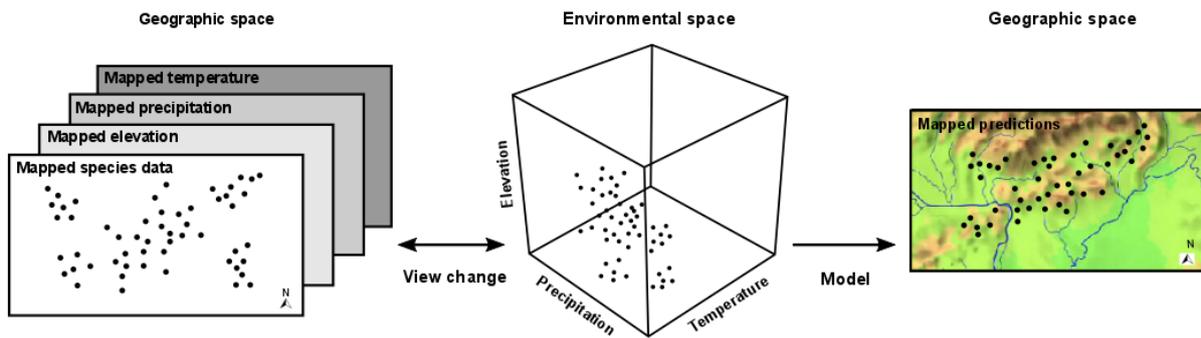


Figure 5.2.1 Construction of abundance based ecological tolerance and prediction of geographic distribution

Based on maps of species occurrences and environmental variables (GIS database layers, left panel) the presences and absences of the species can be plotted within the space of the environmental factors (middle panel), which serves as the input of correlative models predicting the (potential) occurrences of the species at other locations. Given the maps of the same environmental factors for the new area, the predicted map of presences and absences (right panel) can be constructed.

The common principle of different correlative methods is the *habitat matching rule* (Cassini 2011), which simply states that the more often we find a species in a certain environment, the more probable that it is suitable for the species (i.e., $r_i > 0$). Let us consider Table 5.2 (p.80) with this rule in mind! Both the errors of missed presence and fallacious presence/absence may be seriously misleading, if these errors do not occur at random within the space of environmental variables, because they will predict that certain environments are more (or less) suitable for the target species than they really are. This problem may show up quite often in practice, because fallacious presence and absence are often topographically aggregated, and environmental factors are often spatially auto-correlated, too. We may expect accurate forecasts only if the habitat matching rule is met. The problem is illustrated by the study of Duncan et al. (2009) on the Australian and South-African distribution of five dung beetle species. All of these species are native to South Africa; African distribution data come from museum collections. The beetles were deliberately introduced to Australia to clean cow dung, and their populations were monitored. Presence through a couple of years at the location of release was considered as an occurrence, and failed attempts of introductions as absences, i.e., the data were presence/absence type. The model built on the Australian data base has proven to be of excellent predictive value in Australia, because there were no fallacious presences/absences and missed presences of the dataset. The same model based on South-African museum data loaded with these types of errors, performed much worse in predicting either the South-African or the Australian distribution of the species (Figure 5.2.2).

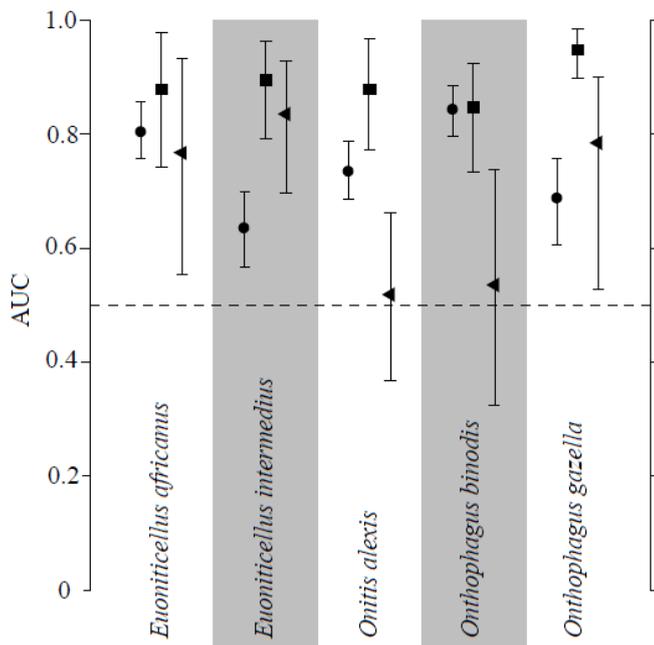


Figure 5.2.2. Prediction accuracy and data quality

Duncan et al. (2009) have built a species distribution model based on Australian and South African data on 5 species of dung beetle (details in main text). They accuracy of predictions were tested using independent data (i.e. not used in model building) with a measure called [AUC](#) (area under receiver operation curve). By randomly choosing one location from the detected occurrences and one from the detected absences the AUC value is the probability that the prediction is correct. AUC=1 indicates that the model is the best possible; AUC=0.5 corresponds to the goodness of a model in which presences and absences are randomly allocated to each sampling point. When the high quality experimental data from Australia were used for predicting occurrences in Australia the predictions were highly accurate (squares). However, South-African data provided worse predictions for South-African occurrences (circles) and even weaker ones when the model was extrapolated to Australian occurrences (triangles). The evaluation process based on random choices was repeated several times thus the mean AUC values and their 95% confidence intervals are also given for the investigated five species (after Duncan et al. 2009).

The spatial or temporal extrapolations of the results of correlative distribution models are hardly ever reliable, because a host of criteria regarding ecological tolerance, the correlation of different environmental variables, and the speed of migration should be met for the extrapolated model to work under circumstances different from those producing the original pattern of distribution. Temporal extrapolation can be made under the assumption that the tolerance of the species does not change within the time range involved. Spatial extrapolation requires that the population at a different location has the same tolerance as the reference population (i.e., there are no ecotypes). Furthermore, it is assumed that the environmental variables omitted in the reference study can be safely disregarded also in the new location or within the new time frame. For example, if the model does not account for the presence of the food plant for an insect species as it occurs everywhere within the study region, then by extrapolating the model we automatically assume that the same condition applies somewhere or sometime else, too.

In general, correlative models tend to disregard regulating (both biotic and abiotic) environmental factors, because it is difficult to obtain the corresponding data. Fortunately, resource levels are often correlated to the environmental variables used in the models. Climatic effects may influence the resource level or the density of natural enemies of the

target population, not the target population itself. The presence or absence of competitors also influences the level of the regulating variables which may lead to competitive exclusion or release. Anderson et al. (2002) investigated this competitive exclusion hypothesis by modelling the distribution of two spiny pocket mouse (*Heteromys* sp.) species on the basis of occurrence data (Figure 5.2.3). These two closely related species belong to the exclusively neotropical pocket mice (*Heteromyidae*) family (Rogers and Gonzalez 2010). Their diets composed of seeds mainly. The members of the family are all thought to be habitat specialists to such a degree that their species number is considered as an indicator of habitat diversity on large scale (Fernandez et al. 2014). The members of the species rich *Heteromys* genus also show strong habitat preference, thus may have evolved by spatial niche segregation (Ch10.2.6, p.222) or in geographical isolation. Of the two species investigated *H. australis* seems to be of higher habitat specificity as it has been collected mainly in humid evergreen forests while *H. anomalus* occurs in deciduous and evergreen tropical forests as well.

Anderson's analysis found two environmental variables: elevation and mean annual precipitation to explain the large-scale geographic distribution of the two species. The typical habitat of *H. australis* is at higher altitudes with cooler climate and more precipitation; *H. anomalus* occurs in dryer forests at lower altitudes (Figure 5.2.4). The ecological tolerance ranges of the two species predicted by a correlative model, based solely on these modifying environmental variables, showed considerable overlap: 38% of the predicted ecological tolerance range of *H. australis* and 18% of *H. anomalus* belonged to the overlapping tolerance range (Tbox 5.1. p.78). From the 21 data collection sites that fall into the region of tolerance range overlap, 17 are inhabited by *H. australis* while 4 are inhabited by *H. anomalus* alone. The 4 locations of potential tolerance range overlap inhabited by *H. anomalus* are in distant and isolated mountains farther away from the rest of the Andes, probably inaccessible for *H. australis* migrants through lower elevations.

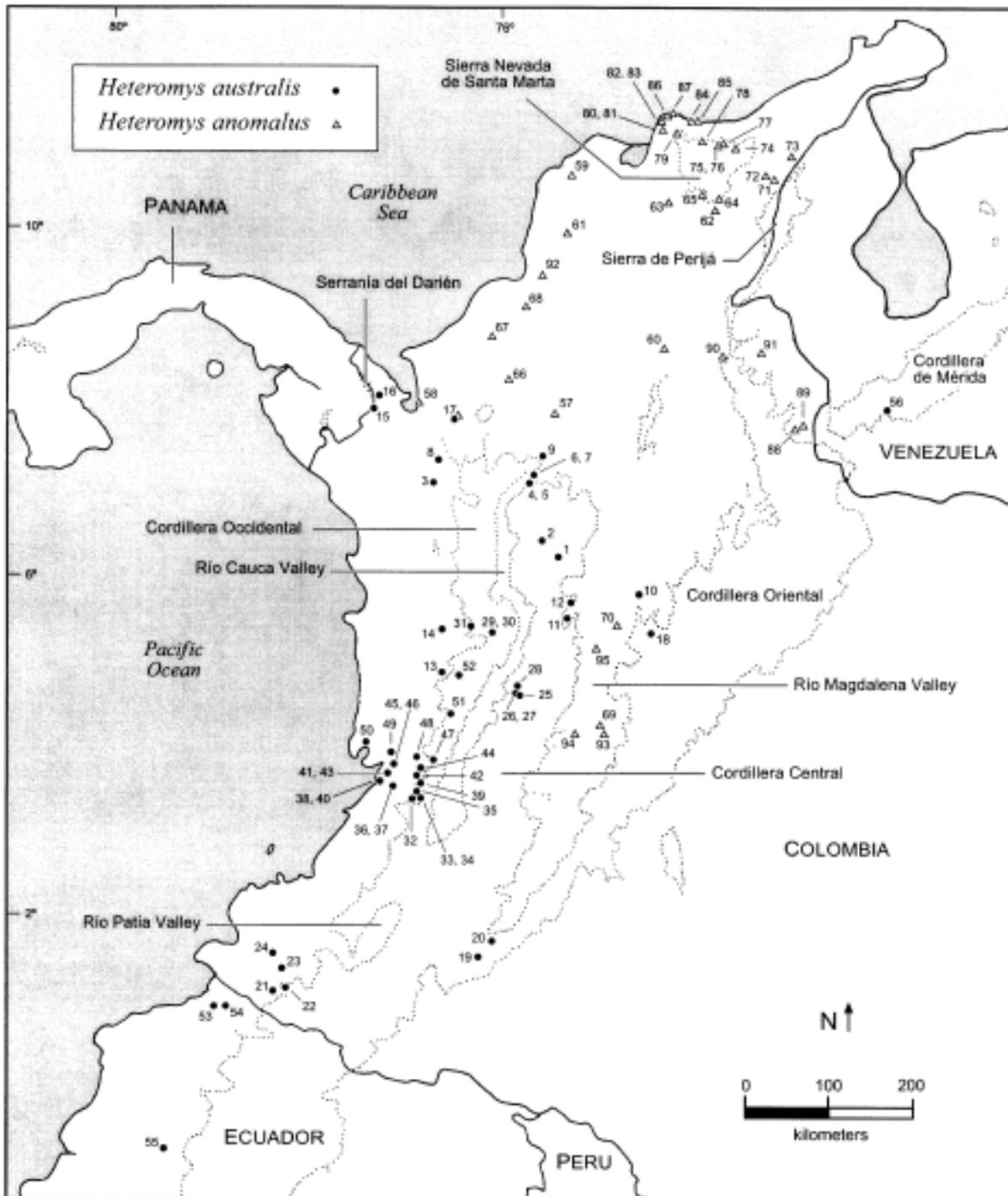


Figure 5.2.3: Collection locations of two spiny pocket mouse (*Heteromys*) species. The map shows collection localities used in the modelling study of the joint distribution of *H. australis* and *H. anomalus* by Anderson et al. (2002). Dotted lines denote regions over 1500 m (Anderson et al. 2002). Reprinted by permission of John Wiley & Sons.

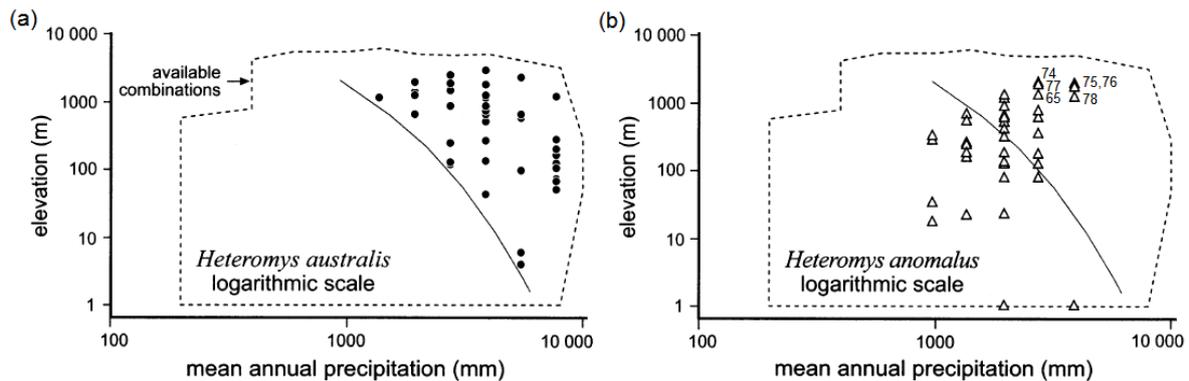


Figure 5.2.4. Abundance based tolerance ranges of two spiny pocketmouse (*Heteromys*) species. The two modifying environmental factors (precipitation and elevation) have proven to be the most relevant for explaining geographical distribution according to the model. The continuous line is the border of the tolerance space of *Heteromys australis* on both panels. On the right panel the localities of *H. anomalus* farthest into the tolerance space of *H. australis* represent sites in the Sierra Nevada de Santa Marta (Localities 74–76, 78 on the map of Figure 5.2.3) and two nearby sites (Localities 65 and 77) where *H. australis* is not present (after Anderson et al. 2002). Reprinted by permission of John Wiley & Sons.

Anderson et al. (2002) interpreted this result in terms of competitive exclusion and release. The two species usually do not occur in sympatry, they have only a narrow contact zone (Figure 5.2.3). Their estimated ecological tolerance ranges are different but overlap (Figure 5.2.4), so there are some geographical region where competitive exclusion must have occurred. If competitive exclusion occurs, the inferior species must show wider abundance based ecological tolerance in isolation (*fundamental ecological tolerance*) than in those regions which are accessible to both species (*competitive release*). According to the analysis of Anderson et al. this is the actual case. *H. australis* is the dominant competitor while *H. anomalus* is the inferior one. Disregarding the actual presence of the competitor leads to under estimation of the tolerance range of the inferior species by the abundance-based method. On the other hand, disregarding the potential presence of the competitor would yield a prediction for the potential geographic range of the inferior species (*H. anomalus*) much larger than found in nature. That is, taking the levels of the regulating environmental factors or the potential presence of competitors into consideration are required for an accurate estimation of ecological tolerance and prediction of geographic distribution. These problems are receiving increasing attention and they are addressed for example, as choices between “Eltonian and Grinnellian models” (Manuel Alvarez-Martinez et al. 2015).

The actual population dynamical effects of different environmental variables are seldom independent of each other; therefore, predictions based on new combinations of environmental effects not used in model calibration may be dubious. On the other hand, variables not directly affecting the species may be good predictors of its occurrence, due to correlation between such variables and the effective environmental factors. Extrapolations

in such cases will be reliable only if the correlation between the proxy and the regulating variable is preserved throughout the range of extrapolation.

Predicting the effects of climate change on species ranges would require taking dispersal abilities of the target species into consideration in the distribution model applied, but data on dispersal are rarely available. The usual workaround is to assume the two extremes in such cases: either no dispersal/migration at all (whereby climate change can only decrease species range – over-pessimistic estimate) or full migration (over-optimistic estimate) allowing for instantaneous range expansion to any location that becomes environmentally suitable due to climate change. The results of these extreme assumptions might be very different: the predicted average change in European herptile (reptile and amphibian) community composition is four times larger with the full migration assumption than it is for no migration (Thuiller et al. 2008).

Since the criteria of extrapolation are rarely met, extrapolated results need to be treated with caution even if the model performs well in interpolation studies. Beaumont et al. (2009) applied eight different statistical procedures to predict the locations potentially invaded by European hawkweed (*Hieracium*) species on other continents. Using calibration data only from the native range of orange hawkweed *Hieracium aurantiacum* (i.e., extrapolating the model predictions in space), 64% of known invasion occurrences were not predicted by any one of the eight methods. Using additional calibration data from all known invasion occurrences increased the accuracy of the model dramatically: each occurrence was predicted by at least two statistical procedures, and 78% of the occurrences were predicted by all the eight of them. There was no substantial difference between the two predictions regarding invasions within Europe; i.e., spatial interpolation was accurate in both cases.

In spite of the limitations discussed above, correlative distribution models are useful tools of ecological research, especially if we keep in mind these limitations when we design the analysis and interpret the results.

References

- Anderson, R. P., A. T. Peterson & M. Gómez-Laverde (2002) Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. *Oikos*, 98, 3-16.
- Aragón, P., A. Baselga & J. M. Lobo (2010) Global estimation of invasion risk zones for the western corn rootworm *Diabrotica virgifera virgifera*: integrating distribution models and physiological thresholds to assess climatic favourability. *Journal of Applied Ecology*, 47, 1026-1035.
- Beaumont, L. J., R. V. Gallagher, W. Thuiller, P. O. Downey, M. R. Leishman & L. Hughes (2009) Different climatic envelopes among invasive populations may lead to underestimations of current and future biological invasions. *Diversity and Distributions*, 15, 409-420.
- Cassini, M. H. (2011) Ecological principles of species distribution models: the habitat matching rule. *Journal of Biogeography*, 38, 2057-2065.
- Czúcz, B., L. Gálhidy & C. Mátyás (2011) Present and forecasted xeric climatic limits of beech and sessile oak distribution at low altitudes in Central Europe. *Annals of Forest Science*, 68, 99-108.
- Dullinger, S., I. Kleinbauer, J. Peterseil, M. Smolik & F. Essl (2009) Niche based distribution modelling of an invasive alien plant: effects of population status, propagule pressure and invasion history. *Biological Invasions*, 11, 2401-2414.
- Duncan, R. P., P. Cassey & T. M. Blackburn (2009) Do climate envelope models transfer? A manipulative test using dung beetle introductions. *Proc Biol Sci*, 276, 1449-57.
- Elith, J. & J. R. Leathwick (2009) Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, 40, 677-697.
- Fernandez, J. A., M. S. Hafner, D. J. Hafner & F. A. Cervantes (2014) Conservation status of rodents of the families Geomyidae and Heteromyidae of Mexico. *Revista Mexicana De Biodiversidad*, 85, 576-588.
- Manuel Alvarez-Martinez, J., S. Suarez-Seoane, C. Palacin, J. Sanz & J. C. Alonso (2015) Can Eltonian processes explain species distributions at large scale? A case study with Great Bustard (*Otis tarda*). *Diversity and Distributions*, 21, 123-138.
- Peterson, A. T. (2006) Uses and Requirements of Ecological Niche Models and Related Distributional Models. *Biodiversity Informatics*, 3, 59-72.
- Raxworthy, C. J., E. Martinez-Meyer, N. Horning, R. A. Nussbaum, G. E. Schneider, M. A. Ortega-Huerta & A. Townsend Peterson (2003) Predicting distributions of known and unknown reptile species in Madagascar. *Nature*, 426, 837-841.
- Rogers, D. S. & M. W. Gonzalez (2010) Phylogenetic relationships among spiny pocket mice (*Heteromys*) inferred from mitochondrial and nuclear sequence data. *Journal of Mammalogy*, 91, 914-930.
- Thuiller, W., C. Albert, M. B. Araújo, P. M. Berry, M. Cabeza, A. Guisan, T. Hickler, G. F. Midgley, J. Paterson, F. M. Schurr, M. T. Sykes & N. E. Zimmermann (2008) Predicting global change impacts on plant species' distributions: Future challenges. *Perspectives in Plant Ecology, Evolution and Systematics*, 9, 137-152.
- Thuiller, W., S. Lavergne, C. Roquet, I. Boulangeat, B. Lafourcade & M. B. Araujo (2011) Consequences of climate change on the tree of life in Europe. *Nature*, 470, 531-534.