ENVIRONMENTAL-ENVELOPE BASED HABITAT-SUITABILITY MODELS

ALEXANDRE H. HIRZEL[1] and RAPHAËL ARLETTAZ, Institute of Zoology, Division of Conservation Biology, University of Bern, CH-3012 Bern, Switzerland.

Abstract: Although all habitat-suitability modelling methods are inherently linked to the Hutchinsonian definition of the ecological niche, those based on the environmental envelopes are its direct geometrical application. Their common principle is to model in the environmental space, a hyper-surface, the envelope, encompassing all or part of the observation points. This paper compares four such methods: the first method based on the distance to the closest observation, or minimum distance. The second based on the frequency distribution of the observations on each dimension. The third and fourth are based respectively on the geometric and harmonic means of the distances to all observation points. We applied these four methods to the case of the Bearded Vulture (*Gypaetus barbatus*) in the Swiss Alps. An Ecological Niche Factor Analysis (ENFA) summarised the 14 environmental variables related to this bird into seven uncorrelated factors, which uses the dimensions of the environmental space. Eventually, a *k*-fold cross-validation process allowed us to compare and discuss these methods and their domain of application.

Key words: Environmental distance, geometric mean, harmonic mean, Ecological Niche Factor Analysis (ENFA), bearded Vulture, *Gypaetus barbatus*.

INTRODUCTION

All habitat suitability modelling methods (for a review, see Guisan and Zimmermann 2000) are inherently based on the Hutchinsonian concept of ecological niche (Hutchinson 1957). Among these methods, those based on the environmental envelopes are conceptually the closest to the niche theory, as they strive to delineate the hyper-surface (or envelope) that best circumscribes the suitable conditions in the space defined by the environmental variables. In a one-dimensional environmental space (e.g. altitude), it comes down to finding the range of values that allow the species to persist, (e.g. from 400 to 700 meters above sea level). In a multi-dimensional space, the problem becomes more complex because of possible interactions among the variables. The basic idea is that the environmental-space coordinates of the sites where we recorded focal species are representative of suitable variable combinations. Intuitively, observed density in this space was correlated to habitat suitability. All environmental-envelope methods are based on this assumption.

Several methods developed on these principles show the increase in computer power allowing ecologists to include more and more details. For instance, BIOCLIM (Busby 1991), a method that was widely used in Australia (e.g. Busby 1988, Carpenter et al. 1993), simply bounds the species occurrences inside hyper-boxes. However, this approach obviously suffers from its coarseness when the species distribution is influenced by a combination of the environmental predictors rather than by each one individually (Carpenter et al. 1993). Walker and Cocks (1991) addressed this issue in the program HABITAT, a complex algorithm based on multidimensional convex sub-envelopes tightly fitted to the data, the environmental predictors being selected through a classification and regression tree (CART) approach. Both methods are binary classifiers, sorting the sites as suitable or unsuitable (nevertheless, core and marginal envelopes can be delineated). Carpenter et al. (1993), comparing the two precedent methods, showed that BIOCLIM tended to give too "permissive" predictions while HABITAT was far too restrictive; they proposed a new method, named DOMAIN, which computes for all locations of the studied area a similarity coefficient based on the Gower's metric (Gower 1971) (eq. 1):

$$d_{AB} = \tfrac{1}{p} \sum_{i=1}^{p} \frac{|A_i - B_i|}{\text{range}_i} \qquad (1)$$

where d_{AB} is the Gower's distance between the points A and B and p is the number of dimensions in the environmental space. This distance is thus simply the arithmetic mean of the differences between the two points in each dimension, this difference being range-standardised to equalise the contribution from each predictor. The similarity coefficient is the inverse of the Gower's distance to the closest species point. Quantitative mapping of the habitat suitability computes this coefficient for all locations of the study area. The main drawback of DOMAIN is that it does not consider the density (in the environmental space) of the observations, thus risking to be misled by outliers.

Hirzel et al. (2002a) took into account the observation density in the "median" algorithm (BIOMAPPER 2.0, Hirzel et al. 2002b). In this method, the environmental predictors are first summarised into a few uncorrelated factors by the Ecological Niche Factor Analysis, a procedure similar to the Principal Component Analysis but producing ecologically meaningful axes that define the space that delineates the environmental envelope (Appendix A). He then computed frequency distribution and median of the record sites on each factor. To compute the habitat suitability of any location, one must compare it to these distributions: the farther the location is from the medians, the lowest its suitability (Figure 1). This process is similar to DOMAIN except that the similarity coefficient depends on the density

---

[1]E-mail: Alexandre.Hirzel@nat.unibe.ch

of recorded sites in the environmental space rather than on the distance to each individual point. Nevertheless, this algorithm assumes that the factor distributions are unimodal and symmetrical. Although this assumption is generally correct, there are exceptions. For instance, asymmetry in the environmental space may arise when the study area is located in a marginal part of the species range, where the most suitable environmental conditions are scarce and where the species has to settle down into sub-optimal areas. The median algorithm will over-evaluate these regions, whereas under-evaluating the most suitable ones. Multimodality is probably more rare but may be caused by (1) historical events, a spreading species settling down in a sub-optimal area before it discovers a better one; (2) interaction with a more competitive species that pushes it away from its intrinsic optimum; or (3) by adaptation to two or more kinds of environment. As the median lies somewhere in between the optima, this algorithm produces wrong results.

In this study, we compared four algorithms: One, the median algorithm (Hirzel et al. 2002a), is based on the frequency distribution of the observations in the environmental space; the three others, geometric mean, harmonic mean and minimum distance, are functions of the distances to the observation points in the environmental space. We implemented these algorithms into BIOMAPPER 2.0, which are available for downloading at the following address: http://www.unil.ch/biomapper. A systematic guide explaining how to replicate these analyses is also at this address.

We applied these algorithms to model the habitat suitability of the Bearded Vulture (*Gypaetus barbatus*) in Valais (Swiss Alps). This species offers a nice example of a multimodal distribution, caused by its reintroduction history.

METHODS AND DATA

Environmental space

With the GIS databases presently available, the environmental space commonly consists of many variously inter-correlated environmental variables. They typically use several different units and scales. Therefore, in order to equalise the contribution of each variable, they must be standardised beforehand. There are various ways to do this (e.g. mean-SD standardisation or Principal Component Analysis (PCA) factorisation). We used the Ecological Niche Factor Analysis (ENFA). This analysis, similarly to the PCA, summarises all information into a few uncorrelated and standardised factors (Hirzel et al. 2002a). The first factor explains all the marginality of the species, i.e. how it differs from the average conditions of the study area; if the species' habitat distribution is bimodal or asymmetrical, this will be reflected by this factor. The other factors explain the species specialisation, i.e. how selective it is by comparison to the available range of environmental conditions; species distribution on these factors is fairly symmetrical and unimodal (These concepts are further explained in appendix A, as well as in Hirzel et al. 2002a). A few factors usually explain the major part of the information, and constitute the dimensions of the environmental space. Moreover, the amount of information explained by each factor weights the environmental space dimensions in the habitat suitability algorithms.

Median algorithm

To compute the median algorithm, we divided the species range on each dimension in a series of classes, in such a way that the median would exactly separate two classes. For every point in the environmental space, we count the number of observations that are either in the same class or in any class farther apart from the median (Figure 1). To achieve normalisation, twice this number divides the total number of observations. Thus, a point belonging to one of the two classes directly adjacent to the median will get a value of one, and a point lying outside the observation distribution will get a value of zero. Lastly, the overall suitability index for this point is computed by the weighted average of its scores on each dimension, the weights being given by the amount of information explained by each dimension.
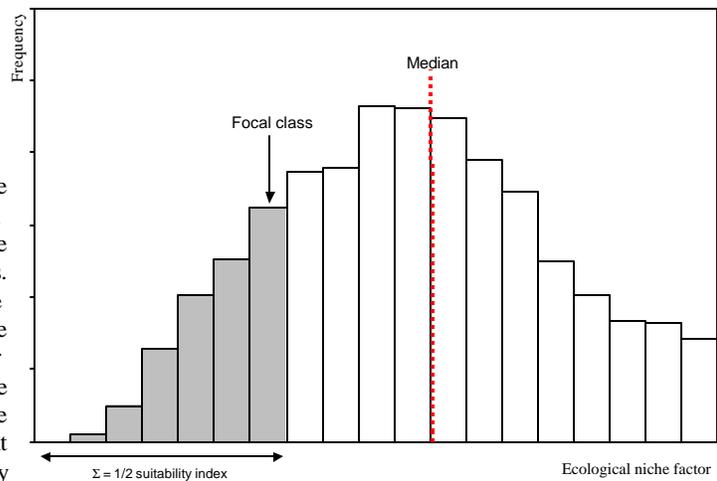
Figure 1: In the "median" algorithm, the suitability of any point in the environmental space computed from its location (arrow) is relative to the species distribution (histogram) on all retained ecological niche factors. Specifically, it is twice the shaded area (sum of all cells from the species distribution that lay as far or farther from the median, dashed vertical line) divided by the total number of cells from the species distribution (surface of the histogram).

Distance geometric mean algorithm

The principle of this method is to draw in the environmental space the influence field, or *suitability field*, of each species observation point in such a way that, when they are close together, their attraction powers reinforce each other.

For any point in the environmental space, one computes the geometric mean of the distances to all observations (equation 3a). Thus, the denser the species points in the environmental space, the higher the habitat suitability. This actually comes down to calibrating a model in the environmental space to apply it to the geographic space. This process does not need absence points, which are most difficult to collect and to rely on (Kéry 2002, Zaniewski et al. 2002).

Let O_i be the N species observation points given by their coordinates in the D-dimensional environmental space. In this space, the weighted Euclidian distance between two points A and B, whose coordinates are $A_1, A_2,…, A_D$ and $B_1, B_2,…, B_D$ respectively, is given by the equation 2:

$$d(\mathbf{A}, \mathbf{B}) = \sqrt{\sum_{i=1}^{D}(B_i - A_i)^2} \qquad (2)$$

The habitat suitability of any point P of the environmental space is the geometric mean H_G of all observation points:

$$H_G(\mathbf{P}) = \sqrt[N]{\prod_{i=1}^{N} d(\mathbf{P}, \mathbf{O}_i)} \qquad (3a)$$

These means range from zero to infinity.

Other distance-based algorithms

Alternatively, functions other than the geometric mean are possible. We chose here to study the Harmonic mean H_H (equation 3b) and the Minimum H_M (equation 3c) functions.

$$H_H(\mathbf{P}) = \cfrac{1}{\cfrac{1}{N}\sum_{\substack{i=1 \\ \mathbf{P} \neq \mathbf{O_i}}}^{N} \cfrac{1}{d(\mathbf{P}, \mathbf{O}_i)}} \qquad (3b)$$

$$H_M(\mathbf{P}) = \operatorname*{Min}_{i=1}^{N}\left\{d(\mathbf{P}, \mathbf{O}_i)\right\} \qquad (3c)$$

The distance harmonic mean is commonly used to delineate home ranges and activity centres from detection locations (Dixon and Chapman 1984), in the geographic space. Here we use it in the N-dimensional environmental space.

By contrast, the minimum distance does not take into account observation density; only the distance to the closest observation is considered, making this algorithm quite similar to DOMAIN (Carpenter et al. 1993).

Envelope modelling

These four methods compute a suitability field, i.e. they attribute a different habitat suitability value to any point in the environmental space. In this space, an envelope defined as the hyper-volume circumscribes all values that are lower than an arbitrarily fixed threshold. Computing an envelope's habitat suitability index occurs by counting the fraction of observation points it contains. We define a core envelope (containing 50% of the observations), a marginal envelope (from 50% to 90%) and an unsuitable domain (the remaining space). These envelopes convert into the corresponding areas in the geographic space. These thresholds were convenient for this study; however, depending on the focal species, one might consider using other thresholds or more envelopes.

Studied species

To illustrate these methods we present an application to the Bearded Vulture in the Valais (Alps, Switzerland); it is, however, beyond the scope of this paper to fully exploit the extracted ecological information from such analyses. The interested reader will find them in another paper fully dedicated to this aspect (Hirzel and Arlettaz submitted).

Since 1986, the reintroduction of 114 individual Bearded Vultures into the Alps (Austria, Switzerland, France and Italy) occurred. In the French Alps (Haute-Savoie), the first releases took place in 1987 at a reintroduction site located close to the Swiss border (Valais). Throughout the Alps, ornithologists, both professionals and amateurs, have been monitoring birds' movements since the beginning of the release programme. In Valais, a network of observers (Réseau Gypaète Valais) have collected and checked 1560 sightings, which stem from at least 29 different individual birds, from 1987 to 31 December 2001. Most identified birds (19 out of 29, ca. 65%) originated from the release site in Haute-Savoie. Data recorded include, among others, date of observation and geographic location. We controlled for any possible bias in observations' clustering by subdividing the study area into major valley systems for which observational effort within a Golden Eagle (*Aquila chrysaetos*) monitoring programme had been quantified (P.A. Ogier) (Figure 2). Recorded Eagles and Bearded Vultures sightings from this study make it possible to estimate a frequency of observations per invested observation time unit for each zone. This enabled us to assess that the observed

clusters of vulture sightings were reflecting real ecological patterns. Moreover, we tested the sensitivity of our models to observations' clustering by sub-sampling the whole data set as follows: whenever several sightings occurred in a same square kilometre in a given year, only one was included into the analysis. The results proved similar to those produced by the full dataset, even tending to be slightly less accurate. This indicated that the additional information provided by multiple sightings per year in a same location was relevant.

The colonisation of Valais showed two chronological phases. The first is a prospective phase (1987-1994), where immature individuals most often visited the southwestern parts of Valais, i.e. the valleys south of the Rhone, in the periphery of the French release site (Figure 2). The second is an installation phase (1995-2001), where maturing birds (subadults) attempted to settle down in the northwestern part of Valais, in calcareous mountain ridges.

Study area

The study area was the territory of the canton of Valais, in the Swiss Alps. It covers 5,191 km^2 modelled by a 100-by-100-meter resolution raster map, i.e. 519,124 grid cells. We used 14 environmental variables, which represented topographic, geologic, biological and anthropogenic information (listed in table 1). To take into account the exploration behaviour of the vulture, included are integrated variables on a 2-km radius. We also tested a shorter (500-m) and longer (infinite) radius, but as they provided consistently lower niche factor coefficients, we chose to discard them from the present study. Whenever it was relevant, these variables were submitted to the Box-Cox standardising algorithm (Sokal and Rohlf 1981) to make their distributions more symmetrical.
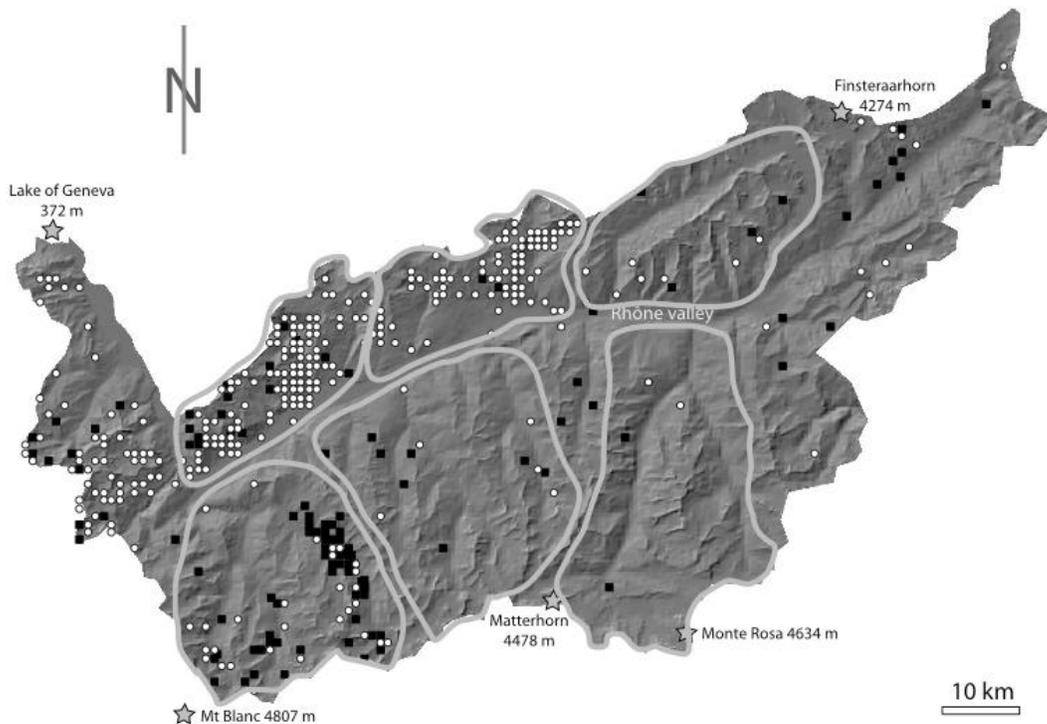
Figure 2 Map of the study area (Valais, Switzerland) showing Bearded Vulture observations made from 1987 to 1994 (immature birds, white circles) and from 1995 to 2001 (subadults, black squares). We controlled for independency between observation density and search effort among the grey-delineated regions.

Algorithm evaluation

The lack of absence data makes it difficult to assess the predictive power of a presence-only based model. Our approach was adapted from the *k*-fold cross-validation proposed by Boyce et al (2002). Of the 10 subsets portioned from the observation data set, nine calibrated a model and the last one validated it. Repeating this step ten times, each time excluding a different testing subset, provided a distribution of the validation results and a way to assess the robustness and the prediction power of the evaluated algorithm. We evaluated the model performances by looking at the predicted suitability indices for the testing subsets. The three envelopes, namely core, marginal and unsuitable habitat, defined three bins. For each bin, we computed the area-adjusted frequency of its cross-validation points; area-adjusted frequencies are the frequency of testing points lying within a bin, divided by the frequency of locations belonging to that bin across the study area. An area-adjusted frequency of 1.0 therefore indicates that cross-validated

testing points occur at rates expected by chance. The area-adjusted frequencies of a model with good predictive power are expected to be high for the core envelope and lower than one for unsuitable habitat (Boyce et al. 2002).

Table 1: Environmental variables included into the analyses.

| Variable category	Variable
Topographical	Average elevation in a 2-km radius
	Average slope in a 2-km radius
	SD of elevation in a 2-km radius
	Average northness in a 2-km radius [1]
	Average eastness in a 2-km radius [2]
Geological	Frequency of calcareous area in a 2-km radius
	Rock frequency in 2-km radius
	Scree frequency in a 2-km radius
	Water frequency in a 2-km radius
Anthropogenic	Building frequency in a 2-km radius
Biological	Forest frequency in a 2-km radius
	Ibex biomass
	Ibex frequency in a 2-km radius
	Chamois biomass

[1] Cosine of exposition
[2] Sine of exposition

RESULTS

Ecological Niche Factor Analysis applied to the Bearded vulture

By comparing the eigenvalues to Mac-Arthur's broken-stick distribution (MacArthur 1960, Hirzel et al. 2002a), we retained seven significant factors. The first factor (marginality) showed a preference of the bearded vulture for calcareous areas, with a southward aspect and higher frequency of Ibexes; moreover, elevation was lower than average (global average: 2224 m; species average: 2044 m above sea level) and there was an eastward aspect tendency. The specialisation factors showed that this bird was mostly selective about elevation, rock frequency and calcareous frequency. However, as it is not the scope of the present paper to study bearded vulture ecological requirements, we will not give more details here.
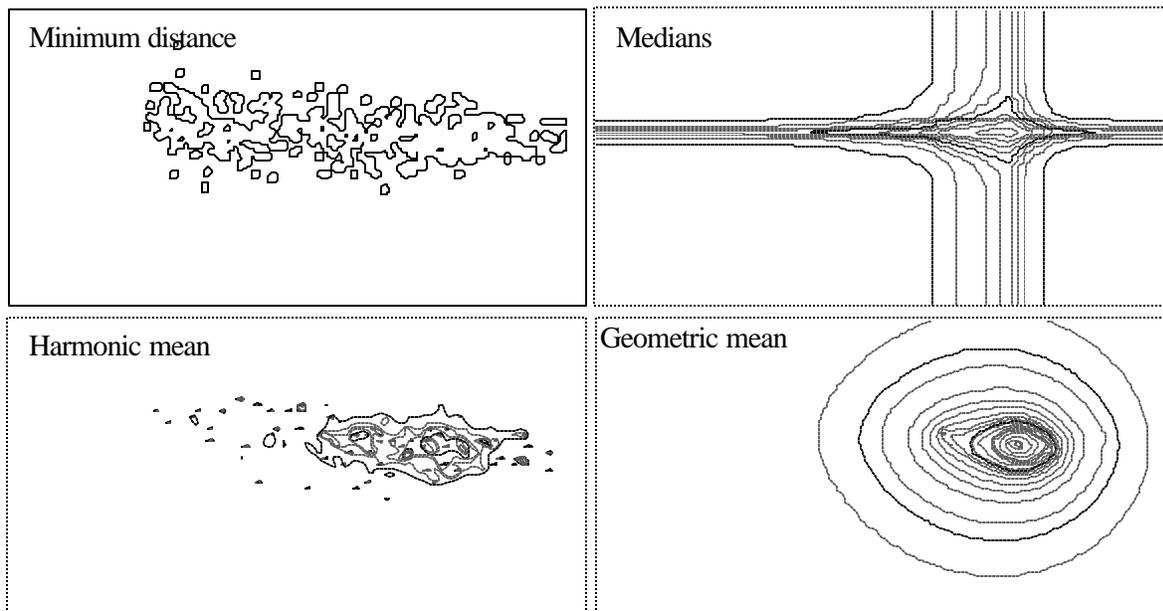
Figure 3: Two-dimensional representation of the environmental space, represented by the two first ecological-niche factors, marginality (horizontal) and specialisation (vertical). The habitat suitability isolines (dashed lines) show how the four algorithms are modelling the niche. Plain lines represent core (bold line) and marginal envelopes. The bimodality of the distribution is obvious in the harmonic mean model.
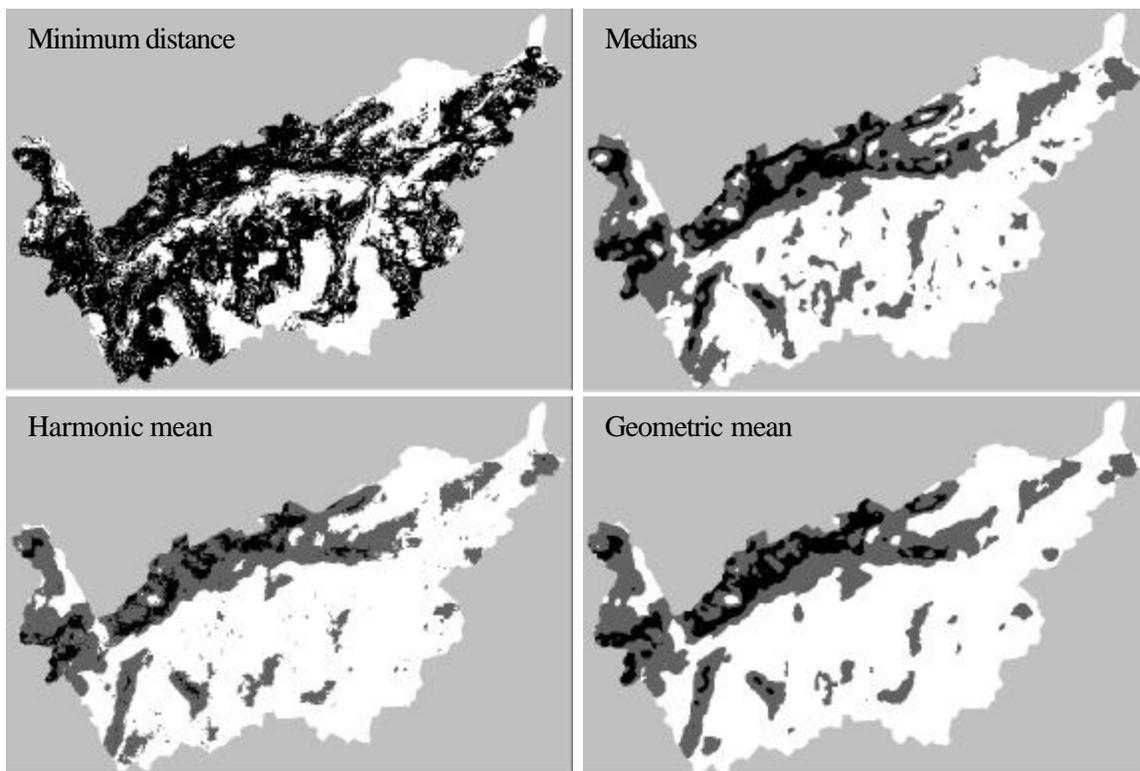
Habitat suitability models

The distribution of species observations in the factorial environmental space has a wide extent, reflecting the far-exploring behaviour of the bearded vulture; however, most sightings occur in two peaks, one broader (the primary patch) than the other (the secondary patch) (Figure 3, harmonic mean).

The four algorithms model this environmental landscape differently. The "minimum distance" algorithm, by construction, can only delineate an envelope encompassing 100% of the observations; consequently, this envelope had a very wide extension. The "median" algorithm located the most suitable habitat on the broadest patch, slightly shifted from its middle point; from this point, the suitability decreased monotonically along a crest in direction of the secondary patch, both of them belonging to the core envelope. Another particularity of this algorithm was the symmetrical cross pattern, extending artificially the marginal habitat quality along the ecological factors. The "harmonic mean" algorithm suitability field was characterised by two main hills surrounded by a constellation of very acute peaks whose sizes decrease with the distance. One or a few observations caused each of these peaks, which was given a heavy weight by the harmonic mean. Finally, the "geometric mean" algorithm identified the broader of the two patches as the core habitat. From this patch, habitat suitability decreased rapidly except in the direction of the secondary patch, where a local maximum of sub-optimal suitability occurred.

Figure 4 shows habitat suitability maps computed by these algorithms. The "minimum distance" algorithm generated a black-and-white map, where good habitat occurred at a much larger extent than in the maps computed by the other methods. The "median" algorithm tended to predict higher suitability values, resulting in a wider marginal habitat; this area was also more fragmented. By contrast, the "geometric mean" predicted fewer and more compact habitat patches. Many small marginal patches predicted by the "median" algorithm were absent from the "geometric mean" map. The map predicted by the "harmonic mean" algorithm was quite similar to the previous one but is more fragmented and shows many tiny spots of marginal habitat.

Figure 5 displays the results of the cross-validation. The worst model was clearly the minimum distance model, although it differs from a purely random prediction. The geometric and harmonic mean models were undistinguishable from this cross-validation point of view, although the maps they generated were slightly different. The median-algorithm core area had the same prediction performance as those generated by the two latter models, but its marginal areas wre not different from random expectation. All four models correctly predicted unsuitable areas.


Figure 4: Habitat suitability maps resulting from the four tested algorithms, showing the spatial distribution of the core (black), marginal (dark grey) and unsuitable habitats (white)

DISCUSSION

The carrion-based diet of the bearded vulture and its great flying capacity allow it to explore daily a very wide area (up to 37 km linear distance from nest site for adults, Arlettaz et al. 2002). One can observe it flying over a great diversity of habitats, many of them having a low quality. Of course, most of the observations will be located close to the best habitats and the successful model must be based on this property and not be too outsider-sensitive. In this respect, the four algorithms gave different results: while the median and geometric mean algorithms produced a good generalisation of the observed density, the harmonic mean and, particularly, the minimum distance algorithms gave more weight to the individual observations. The minimum distance algorithm assigns the same importance to all the observations and would therefore achieve better results on a less vagrant species, whose detections would be more correlated with suitable habitat. Carpenter et al. (1993) and Hargrove & Hoffman (2000), use similar algorithms to achieve good predictions of plant distributions. By contrast, the harmonic mean algorithm uses the observation density information to increase the influence of observations that are closer to the distribution centre. Nevertheless, due to the singularity of the harmonic function on the observations points, they will consistently get a high suitability value. It is this property that is used in applications of this function to delineate the home-range of an animal on the base of radio-telemetry locations (Dixon and Chapman 1984). Although it is counter-productive in the case of the bearded vulture, this property could prove useful for other organisms or when the paucity of the sampling advocates for extracting information from every observation.

Another challenging difficulty of this bird is, in this case, its particular reintroduction history. As stated above, the vultures began by settling down in southwestern Valais before colonising the mountains north of the Rhone. Reflected in the environmental space by a bimodal pattern (Figure 3, harmonic mean) is this shift in the geographic space (Figure 2). The left-hand isolated patch corresponds to the sub-optimal valley that was first colonised and the broad main patch corresponds to the more recently selected area, and the points scattered throughout the environmental space correspond to occasional observations of birds in unsuitable habitats. This particular distribution misleads the "median" algorithm and any model making with too stringent assumptions on the environmental space landscape. However, observation distribution in the environmental space is usually unimodal and grossly symmetric around the optimum, and the median algorithm has been applied successfully to many species (e.g. Hausser 1995, Breitenmoser et al. 1999, Chatelain and Gautier 2000, Dunand 2001, Hirzel et al. 2001, Hirzel et al. 2002a, Zaniewski et al. 2002, Reutter et al. 2003).

Any method using the observation density in the environmental space assumes that this density is the result of an ecological or behavioural process. In this respect, data stemming from observational studies or museum collections
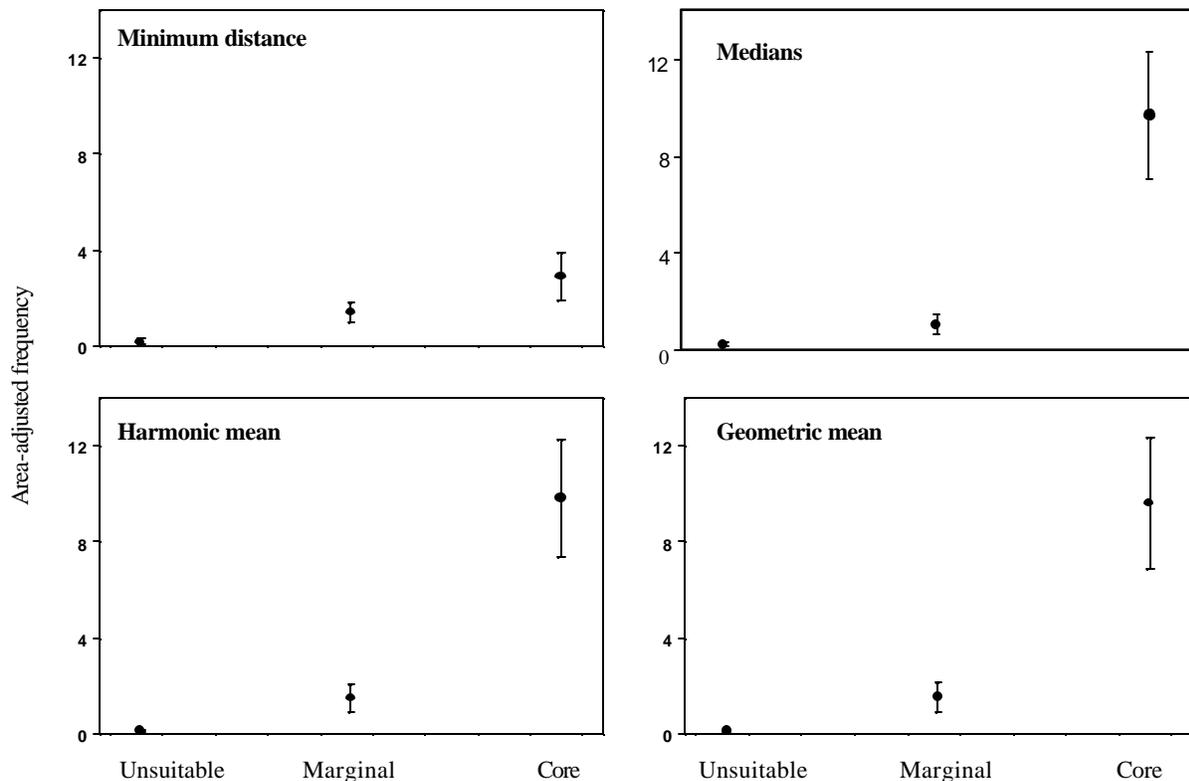
Figure 5: Area-adjusted frequency of cross-validated testing points classed into the unsuitable, marginal and core areas. Values close to one indicate that the observed frequency is not different from random expectation; values smaller than one indicate avoidance and greater than one indicate preference. Black circles represent median and bars represent the 95% confidence intervals.

may suffer from an observation effort bias. Differences in species delectability in different habitats may bias even carefully designed experiments. However, in spite of these problems, we believe that valuable information can still be extracted from such data sets (see e.g. Peterson 2001), providing bias is either accounted for or removed.

The three distance-based algorithms are much slower than the "median" algorithm. While this is not a problem when computing only one model, it can be time-consuming when Monte-Carlo statistics are involved. For example, the k-fold cross-validation procedure we applied here was ten times slower for the distance-based algorithms than for the median algorithm.

Comparing these algorithms to habitat suitability analyses based on other principles was beyond the scope of this paper. However, the median algorithm has already been compared to Generalised Linear Models (GLM, Hirzel et al. 2001, Sachot 2002) and Generalized Additive Models (GAM, Zaniewski et al. 2002). Further tests on extrapolating these analyses to areas outside of the calibrating region is needed.

Sporadicity (i.e. the fact that many observations are just random occurrences and do not convey reliable information about habitat suitability) and multimodality entail contradictory constraints: whereas multimodality needs a close fitting model, sporadicity generates many outsiders, thus ignored by a smoother, more general model. One should therefore select the algorithm on the base of the specificities of the studied organism, of the quality of the available data and on the goal of the model.

ACKNOWLEDGEMENTS

LITERATURE CITED

ARLETTAZ, R., B. POSSE, AND Y. CRETTENAND. 2002. The bearded vulture in Valais in 1999. Pages 58-59 *in* Bearded vulture reintroduction into the Alps, Annual report 2000.

BOYCE, M. S., P. R. VERNIER, S. E. NIELSEN, AND F. K. A. SCHMIEGELOW. 2002. Evaluating resource selection functions. Ecological Modelling 157:281-300.

BREITENMOSER, U., F. ZIMMERMANN, P. OLSSON, A. RYSER, C. ANGST, A. JOBIN, AND C. BREITENMOSER-WÜRSTEN. 1999. Beurteilung des Kantons St.Gallen als Habitat für den Luchs. Expertise KORA, Bern.

BUSBY, J. R. 1988. Potential impacts of climate change on Australia's flora and fauna. Pages 387-398 *in* G. I. Pearman, editor. Greenhouse: Planning for climate change, CSIRO: Melbourne.

BUSBY, J. R. 1991. BIOCLIM - A Bioclimate Analysis and Prediction System. Pages 64-68 *in* C. R. Margules and M. P. Austin, editors. Nature Conservation: Cost effective biologival surveys and data analysis., CSIRO: Melbourne.

CARPENTER, G., A. N. GILLISON, AND J. WINTER. 1993. DOMAIN - a Flexible Modeling Procedure for Mapping Potential Distributions of Plants and Animals. Biodiversity and Conservation 2:667-680.

CHATELAIN, C., AND L. GAUTIER. 2000. Application du SIG IVOIRE à la distribution potentielle des espèces en fonction des facteurs écologiques. *in* Association des Taxonomistes d'Afrique Tropicale, Bruxelles.

DIXON, K. R., AND J. A. CHAPMAN. 1984. Harmonic Mean Measure of Animal Activity Areas. Ecology 61:1040-1044.

DUNAND, I. 2001. Le cerf, un amoureux des p'tites fleurs ? Sélection de l'habitat chez le cerf (Cervus elaphus) au sein d'une population en voie d'expansion dans le massif jurassien. Masters. University of Lausanne, Lausanne

GOWER, J. C. 1971. General Coefficient of Similarity and Some of Its Properties. Biometrics 27:857-874.

GUISAN, A., AND N. E. ZIMMERMANN. 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135:147-186.

HARGROVE, W. W., AND F. M. HOFFMAN. 2000. An analytical assessment tool for predicting changes in a species distribution map following changes in environmental conditions. Pages #5 *in* 4th international conference on integrating GIS and environmental modeling (GIS/EM4): Problems, prospects and research needs., Banff, Alberta, Canada.

HAUSSER, J. 1995. Mammifères de Suisse. Birkhäuser, Bâles.

HIRZEL, A. H., J. HAUSSER, D. CHESSEL, AND N. PERRIN. 2002a. Ecological-niche factor analysis: How to compute habitat- suitability maps without absence data? Ecology 83:2027-2036.

HIRZEL, A. H., J. HAUSSER, AND N. PERRIN. 2002b. Biomapper 2.0. Software. Division of Conservation Biology, Bern. URL: http://www.unil.ch/biomapper

HIRZEL, A. H., V. HELFER, AND F. MÉTRAL. 2001. Assessing habitat-suitability models with a virtual species. Ecological Modelling 145:111-121.

HUTCHINSON, G. E. 1957. Concluding remarks. Pages 415-427 *in* Cold Spring Harbour symposium on quantitative biology.

KÉRY, M. 2002. Inferring the absence of a species: a case study of snakes. Journal of Wildlife Management 66:330-338.

MACARTHUR, R. H. 1960. On the Relative Abundance of Species. American Naturalist 94:25-36.

PETERSON, A. T. 2001. Predicting species' geographic distributions based on ecological niche modeling. Condor 103:599-605.

REUTTER, B. A., V. HELFER, A. H. HIRZEL, AND P. VOGEL. 2003. Modelling habitat-suitability on the base of museum collections: an example with three sympatric *Apodemus* species from the Alps. Journal of Biogeography 30:581-590.

SACHOT, S. 2002. Viability and management of an endangered capercaillie (*Tetrao urogallus*) metapopulation. PhD. University of Lausanne

SOKAL, R. R., AND F. J. ROHLF. 1981. Biometry: The principles and practice of statistics in biological research, 2 edition. W.H. Freeman & co, New York.

WALKER, P. A., AND K. D. COCKS. 1991. HABITAT - a Procedure for Modeling a Disjoint Environmental Envelope for a Plant or Animal Species. Global Ecology and Biogeography Letters 1:108-118.

ZANIEWSKI, A. E., A. LEHMANN, AND J. M. C. OVERTON. 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. Ecological Modelling 157:261-280.

Appendix A: Ecological Niche Factor Analysis

One aim of factor analyses is to transform V correlated variables into the same number of uncorrelated factors. As these factors explain the same amount of total variance, subsequent analyses may be restricted to the few important factors (e.g., those explaining the largest part of the variance) without loosing too much information. The ENFA achieves this by computing ecologically meaningful factors.

Data are in the form of raster maps, which are grids of N isometric cells covering the whole study area. Each cell of a map contains the value of one variable. The eco-geographical maps contain continuous values, measured for each of the V descriptive variables. Species maps contain boolean values (0 or 1), a value of 1 meaning that the presence of the focal species was proved on this cell. A value of zero simply means absence of proof. Alternately, presence can be given a weight reflecting the amount of cell utilization by the focal species.

A point in the multidimensional space of the variables can thus represent each cell. If distributions are multinormal, their distribution will have the shape of a hyper-ellipsoid. The cells where the observed focal species exist constitute a subset of the global distribution and define a smaller hyper-ellipsoid within the global one. The first factor, or *marginality factor*, is the straight line passing through the centroids of the two ellipsoids. The marginality of the species is the standardised distance between these centroids.

Projecting two ellipsoids onto a hyperplane perpendicular to the marginality factor ensures that subsequent factors will be uncorrelated to the first and removes one dimension from the environmental space. Computing the axis that maximizes the ration of global variance on species variance produces the first of the *specialization factors*. Repeating this process extracts the following specialization factors, with each step removing one dimension from the space until extraction of all V factors is complete. The amount of specialization accounted for by a factor is necessarily lower than that due to the one previously extracted. All specialization factors are furthermore orthogonal, in the sense that the distribution of the species subset on any factor is uncorrelated with its distribution on the others. A limited number (F) of the first factors will thus generally contain most of the relevant information. Their small number and independence make them easier to use than the original variables. From this process, it follows that the ecological niche factors depend strongly on the extent of the reference area.

Hirzel et al. (2002) developed the Mathematical demonstration of this procedure.

Appendix references

Hirzel, A.H., J. Hausser, D. Chessel, and N. Perrin. 2002. Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? *Ecology* 83:2027-2036.