Modeling Habitat Suitability for Complex Species Distributions by Environmental-Distance Geometric Mean

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

ABSTRACT / This paper presents a new habitat suitability modeling method whose main properties are as follows: (1) It is based on the density of observation points in the environmental space, which enables it to fit complex distributions (e.g. nongaussian, bimodal, asymmetrical, etc.). (2) This density is modeled by computing the geometric mean to all observation points, which we show to be a good trade-off between goodness of fit and prediction power. (3) It does not need any absence information, which is generally difficult to collect and of dubious reliability. (4) The environmental space is represented either by an expert-selection of standardized variables or the axes of a factor analysis [in this paper we used the Ecological Niche Factor Analysis (ENFA)].We first explain the details of the geometric mean algorithm and then we apply it to the bearded vulture (Gypaetus barbatus) habitat in the Swiss Alps. The results are compared to those obtained by the "median algorithm" and tested by jack-knife cross-validation. We also discuss other related algorithms (BIOCLIM, HABITAT, and DOMAIN). All these analyses were implemented into and performed with the ecology-oriented GIS software BIOMAP-PER 2.0.The results show the geometric mean to perform better than the median algorithm, as it produces a tighter fit to the bimodal distribution of the bearded vulture in the environmental space. However, the "median algorithm" being quicker, it could be preferred when modeling more usual distribution.

Habitat suitability (HS) modeling is inherently based on the Hutchinsonian concept of ecological niche (Hutchinson 1957). All habitat suitability modeling methods (for a review, see Guisan and Zimmermann 2000) try to predict species occurrence on the basis of ecogeographical predictors. Among them, the environmental-envelope based ones are conceptually very close to the niche theory as they try to delineate in the space of the ecogeographical variables—which need to be quantitative—the hypersurface (or envelope) that circumscribes all suitable conditions. In a one-dimensional ecogeographical space, say altitude, it simply means to find the range allowing the species to persist (e.g, from 400 to 700 ms above see level). In a multidimensional space, the problem becomes more complex as the species may be sensitive to particular interactions or a combination of the variables. Practically, the suitable conditions are identified by the position, in the ecogeographical space, of the sites where the focal species has been recorded. Intuitively, occurrence density in this space is correlated to habitat suitability, and all the environmental-envelope methods are based on this property.

As computers became more and more powerful, several methods developed that address this problem with increasing detail. The method BIOCLIM (Busby 1991) simply framed the species occurrences into bounding hyperboxes; it earned a great success and was widely used in Australia (Busby 1988, Carpenter and others 1993). Nevertheless, the bounding-box approach obviously suffers from its coarseness when the species distribution is influenced by a combination of the ecogeographical predictors rather than by each one individually (Carpenter and others 1993). Walker and Cocks (1991) addressed this problem in the program HABITAT, a complex algorithm based on multidimensional convex subenvelopes narrowly fitted around the data, according to ecogeographical variables selected through a classification and regression tree (CART) approach. Both methods are basically binary classifiers, discriminating the sites into suitable or unsuitable (nevertheless, core and marginal envelopes can be delineated).

Carpenter and others (1993) compared the two precedent methods and showed that BIOCLIM tended to be too "permissive" in its predictions while HABITAT was far too restrictive; they proposed a new method, named DOMAIN, which computes for all grid points of

the studied area a similarity coefficient based on the Gower's metric (Gower 1971) (equation 1):

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

where d_{AB} is the Gower's distance between the points A and B, p is the number of dimensions in the ecogeographical space, and range$_i$ is the statistic range of the ith dimension (maximum − minimum). This distance is thus simply the arithmetic mean of the differences between the two points in each dimension, this difference being range-standardized to equalize the contribution from each predictor. The similarity coefficient is the inverse of the Gower's distance to the closest species point. By computing this coefficient for all grid points of the study area, it is possible to map the habitat suitability by a quantitative value. The main drawback of the DOMAIN procedure is that it does not take the density (in the ecogeographical space) of observations into account, exposing itself to being misled by outliers.

Hirzel and others (2002a) took into account the observation density in the software BIOMAPPER 2.0 median algorithm (Hirzel and others 2002b). In this method, the ecogeographical predictors are first summarized into a few uncorrelated factors by the Ecological Niche Factor Analysis (ENFA)—a procedure similar to the Principal Component Analysis (PCA)—which define the space where the environmental envelope will be delineated. The frequency distribution and the median of the observation sites are computed along each factor. Every grid cell can be compared to these distributions and the further the cell is from the medians, the less suitable it is. The habitat suitability index is based on this principle, by computing a weighted average of these frequency-based distances to medians. This process is in principle quite similar to the DOMAIN process except that the similarity coefficient depends on the density of recorded sites in the ecogeographical space rather than on each individual point.

In this paper, we present a new method based on a distance geometric mean metric, which computes the habitat suitability in the environmental space without assuming any distribution of the species points, therefore solving the problems encountered both by the DOMAIN and median algorithms. This new algorithm has been implemented in BIOMAPPER 2.0 (http://www.unil.ch/biomapper) and can therefore be easily compared with the above-mentioned algorithms.

As an illustration, both the median and geometric mean algorithms have been applied to model the habitat suitability of the bearded vulture (*Gypaetus barbatus*) in Valais (Swiss Alps). This species was chosen because it offers a good example of a bimodal distribution due to its reintroduction history.

Data and Methods

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 point in such a way that when they are close together their attraction powers are mutually reinforced. For any point of the environmental space, this is done by computing the geometric mean of the distances to all species points (equation 3 below). Thus, the denser—in the environmental space—the species points around a grid cell, the higher its habitat suitability. Actually we are calibrating a model in the environmental space to apply it in the geographic space. This process does not need absence points, which are most difficult to collect and to rely on (Hirzel and others 2002a, Zaniewski and others 2002).

The dimensions of the environmental space must be standardized beforehand to equalize the contribution of each variable. This can be done in different ways. With BIOCLIM and HABITAT, a small set of variables were selected by an expert and directly fed into the algorithms. The DOMAIN algorithm had to standardize them by dividing each variable by its range. This expert selection and standardization (or a mean-SD standardization) could also be used with the methods we propose here. The variable selection, however, needs to make a prior ecological assumptions about the most important variables and to verify that included variables are mostly uncorrelated. Actually, feeding two correlated variables into such a model decreases the relative weight of the others, potentially biasing the result. As environmental variables are usually correlated, a way to cope with this problem is to transform the variables into uncorrelated factors by means of a factor analysis like PCA.

In this paper, we chose to use the Ecological Niche Factor Analysis (ENFA), which, similar to PCA, summarizes all information into a few uncorrelated and standardized factors but which, unlike PCA, convey ecological information (Hirzel and others 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 specialization, 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 (for a

more detailed description of marginality and specialization factors, see the Appendix in the present paper, and Hirzel and others 2002a).

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 equation 2:

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

where w_i are the weights of the predictors, if any. In our case, the weights are given by the amount of information explained by each ENFA factor.

For any point P of the environmental space, the geometric mean of its distances to all observation points O_i is given by:

$$\mu_G(P) = N\sqrt{\pi_{i=1}^{N}\delta(P,O_i)} \qquad (3)$$

These means range from 0 to infinity. Envelopes can be delineated in this space, encompassing hypervolumes with values below a fixed threshold, i.e., delineating the niche in the environmental space. A habitat suitability index can finally be associated to each envelope by counting the number of observation points they contain, thus defining core areas (e.g., an envelope containing 50% of the points), marginal areas (from 50% to 90%), and unsuitable areas (less than 10% of observations are discarded). Of course, these thresholds are arbitrary and should be discussed carefully in each case, depending on species movement behavior and data quality.

Studied Species

To illustrate this method, we present an application on the bearded vulture in the Valais (Alps, southwestern Switzerland) and compare the results to those obtained by the BIOMAPPER median algorithm. However, it is not in the scope of this paper to fully exploit the species-specific ecological clues that can be drawn from such analyses. The interested reader will find them in a companion paper fully dedicated to this species (Hirzel and Arlettaz, submitted).

Since 1986, the bearded vulture has been reintroduced into the Alps (Austria, Switzerland, France, and Italy). 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 the birds' movements since the beginning of the release program. In Valais, a

network of observers (Réseau Gypaète Valais) has collected and checked 1560 sightings, which stem from at least 29 different individual birds, from 1987 to 31 December 2001. The observers recorded the kilometric cell in which the vulture had been seen first and, if it had transited through several cells, the coordinates of the last one; when this was the case, in order to remove temporal autocorrelation, the average between the two locations was used. In some cases, it was also possible to identify marked birds; however, marking was being obtained by bleaching feathers, and thus it disappears at the first molt, usually after 2–3 years and this information could therefore not be used for our purpose. Nonetheless, we found that most identified birds (19 of 29, ca. 65%) originated from the release site in Haute-Savoie (France).

In an attempt to remove potential dependence among sightings, we tried, in preliminary analyses, to use alternately either only one sighting per spatial location, or only one sighting per year per spatial location; however, models built on this base proved slightly less accurate than using the whole data set, and we present here only those latter results. None of the sightings were made on the same day and same location, ensuring time-independence of the data. Most of the sightings are obviously not independent in the sense that they are due to the same individual; however, this is not a problem for the present methods, which rely only on the frequencies of use.

Although the observation effort was not constant across the whole study area due, in particular, to variable numbers of local observers, we controlled for any possible bias in clustering of observations by subdividing the study area into six major valley systems for which the observational effort in a golden eagle (*Aquila chrysaetos*) monitoring program was quantified (P. A. Oggier) (Figure 1). As the latter eagle specialist systematically recorded bearded vulture sightings in parallel, an index of the frequency of observations per invested observation time unit and year (1990–2001) could be estimated for each zone. These enabled us to assess that clusters of bearded vulture sightings were independent of the observation effort (Hirzel and Arlettaz, submitted). However, this does not entail that absence data could be included to improve the model, as this bird is still in a spreading phase and the sightings are still closer to the release site than expected by chance, although the influence of this variable decreases progressively with time (Hirzel and Arlettaz, submitted). Absence can reliably be used only if the focal species distribution has achieved equilibrium (Hirzel and others 2001). Among the birds released in France from 1987 onwards, several visited Valais soon after the first
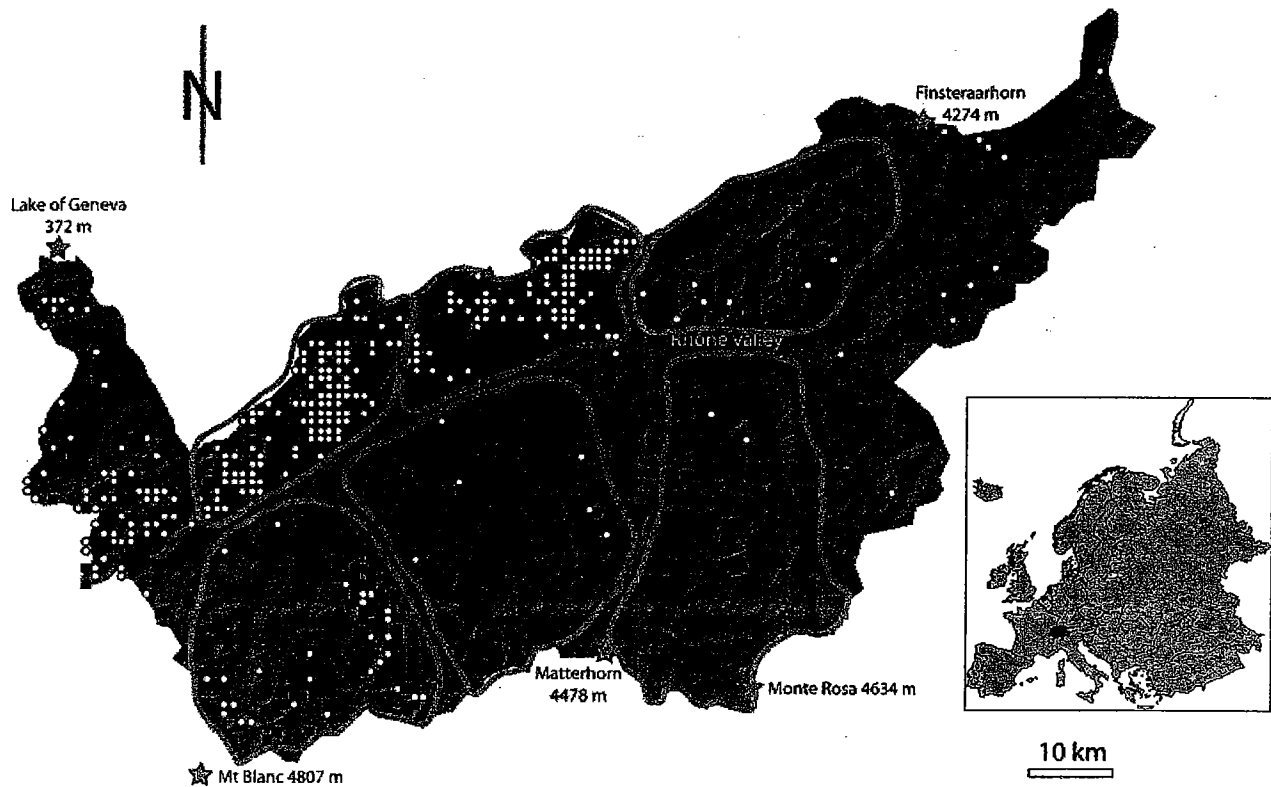
Figure 1. Hill-shade map of the study area (Valais, Switzerland) showing 1-km² squares with bearded vulture sightings from 1984 to 1994 (black squares, prospective phase) and from 1995 to 2001 (white circles, installation phase). No significant relationship was detected between observation effort and sighting number among the six delineated regions. The inset map shows the position of Switzerland inside Europe.

release event, with a few individuals attempting to settle in that area more recently. Accordingly, the colonization of Valais showed two distinct chronological phases: (1) the prospective phase (1987–1994) in which immature individuals visited mostly the southwestern parts of Valais, i.e., the valleys south of the main Rhone valley axis, at the periphery of the release site in France (Figure 1); and (2) the installation phase (1995–2001) in which maturing birds (subadults) attempted to settle in the northwest of Valais, in calcareous mountain ridges.

Study Area

The study area is the territory of the canton of Valais, in the Swiss Alps. It covers 5191 km² modeled by a 100-× 100-m resolution raster map, i.e., 519, 124 grid cells. Fourteen ecogeographical variables were used: topographic, geologic, biological, and anthropogenic (listed in Table 1).

These variables were made more symmetrical by the Box-Cox standardizing algorithm (Sokal and Rohlf 1981). This procedure looks for the best transforma-tion (producing a distribution as close to a gaussian as possible) among the family of functions $Y = (Y^{\lambda} - 1)/\lambda$ by iteratively optimising λ.

Algorithm Evaluation

These data were submitted to ENFA, which extracted mutually independent factors. The significant factors [as determined by comparison to MacArthur's broken-stick distribution (Hirzel and others 2002a)] were used as dimensions of the environmental space for habitat suitability analyses.

Two algorithms were compared on the same dataset: BIOMAPPER's median algorithm (Hirzel and others 2002a) and the above-described distance geometric mean algorithm. They do not directly compute probability estimates but rather assign a similarity coefficient—or habitat suitability index—between any candidate location and the most suitable conditions. These values, combined with the presence points as described above, were used to delineate three regions whose boundaries where arbitrarily fixed, namely core habitat (50% of the presence points), marginal habitat (50% to

Table 1. Ecogeographical variables included in 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[a]
	Average eastness in a 2-km radius[b]
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

[a]Cosine of exposition.
[b]Sine of exposition.

90%), and unsuitable habitat (more than 90%). These thresholds were fixed as the most conforming to expert knowledge. They are, of course, dependent on the studied species. Looking at how validation indices vary for different thresholds may guide the expert decision process.

The lack of absence data makes the predictive power assessment of a presence-only based model difficult. One possibility is to count the proportion of evaluation points occurring in the predicted core habitat. However, according to this absolute validation index method, a model predicting a value of 100 on every cell would be seen, absurdly, as perfect as it would have 100% of the evaluation points above 50. To circumvent this problem when comparing algorithms, we have computed a contrast validation index by comparing the above index to the value that could be expected by chance only: let A_v be the proportion of validation points in the core habitat (HS > 50); by construction of the core envelope, A_v will not, on average, be greater than 0.5. Let A_g be the same proportion for the whole study area. The contrast C value is then given by:

$$C = A_v - A_g \qquad (4)$$

This contrast index ranges from 0 to $(1 - A_g)$, a value of 0 meaning that the model accuracy does not outperform a random model. A high value of the contrast index means that the model is better, both more accurate and more "contrasted"; in presence–absence confusion matrix terminology, this indicates higher sensitivity and specificity (Fielding and Bell 1997). As A_v is generally not greater than 0.5, so is C.

The accuracy of these models was evaluated through jack-knife cross-validation (Sokal and Rohlf 1981, Manly 1991, Fielding and Bell 1997): the data set was partitioned into 100 subsets of which, alternately, 99 were used to calibrate the model (calibration set) and one to evaluate it (evaluation set). This produced the mean and standard deviation on the A_v and C indices, which could then be compared by Student's t tests. The outcome of the t test actually does not depend on the number of partitions as the variance of the validation indices increase with it.

Results

Ecological Niche Factor Analysis Applied to Bearded Vulture

By comparison of the eigenvalues to MacArthur's broken-stick distribution, seven factors were extracted as providing significant information. The first factor (marginality) showed a preference of the bearded vulture for calcareous areas, with a southward exposure (=aspect) and with higher frequency of ibex; moreover elevation is lower than average (global average: 2224 m; species average: 2044 m above sea level), and there is an eastward aspect tendency. The specialization factors show that this bird is mostly selective about elevation, rock frequency, and calcareous frequency. However, as it is not in the scope of the present paper to study bearded vulture ecological requirements, we shall not give here more details.

Habitat Suitability Models

Although the distribution of species observations in the factorial environmental space is widely spread, reflecting the far-ranging behavior of the bearded vulture in the initial prospective phase, most of them are concentrated into three patches, two being closely connected (Figure 2). The isolated peak (Figure 2A) is related to the immature birds whilst the twin peaks (Figure 2B) are due to the subadults. This bimodality is almost completely accounted by the marginality factor as shown by the 1-dimensional histograms represented in Figure 2. This indicates that subadult birds tend to fly over locations with higher frequency of limestone and more southward aspect than the immature do. This environmental landscape is differently modeled by the two algorithms.

The median algorithm locates the most suitable habitat on the middle patch (Figure 2B), the densest of the three, slightly shifted from its middle point, from this point, the suitability decreases monotonically with a crest in direction of the secondary patch, although, at
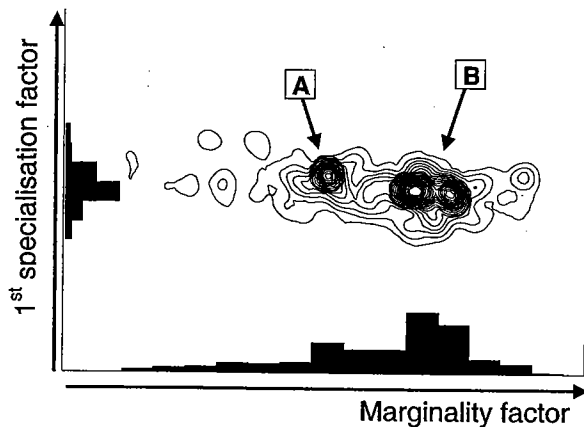
Figure 2. Bearded vulture observation density in the two-dimensional environmental space represented by the two first ecological niche factors, marginality (horizontally) and specialisation (vertically). The density isolines show the bimodality of the distribution; peak A is due to the prospective phase, and the twin peaks B are due to the installation phase. The margin histograms represent the same distribution in one dimension.

this point, the suitability is already quite low (Figure 3A). Another consequence of this algorithm is the cross arms artificially extending the medium habitat quality all along the ecological factors.

The center of the middle patch is identified as the best habitat by the "geometric mean" algorithm with an area of high quality around the twin peaks (Figure 2B). From here, habitat suitability decreases rapidly except toward the secondary patch (Figure 2A) where one observes a local maximum of sub-optimal suitability (Figure 3B).

The habitat suitability maps computed by these algorithms are given in Figure 4. The median algorithm tends to predict higher suitability values, resulting in wider core and marginal habitats. These areas are also more fragmented, with unsuitable patches occurring inside globally good regions. By contrast, the geometric mean predicts fewer and more compact habitat patches. Many small marginal patches predicted by the median algorithm lack in the geometric mean map.

The jack-knife cross-validation computed an average "absolute validation index" of 0.39 (SD = 0.2) for the median algorithm and 0.44 (SD = 0.27) for the geometric mean algorithm; these values differ significantly (t test, $df = 198$, $P = 0.045$). The average contrast validation indexes were, respectively 0.29 (SD = 0.2) and 0.37 (SD = 0.27), which differ very significantly (t test, $df = 198$, $P = 0.01$). These validation measures both tend to indicate that the geometric mean algorithm offers a better generalisation of the data than the median one.
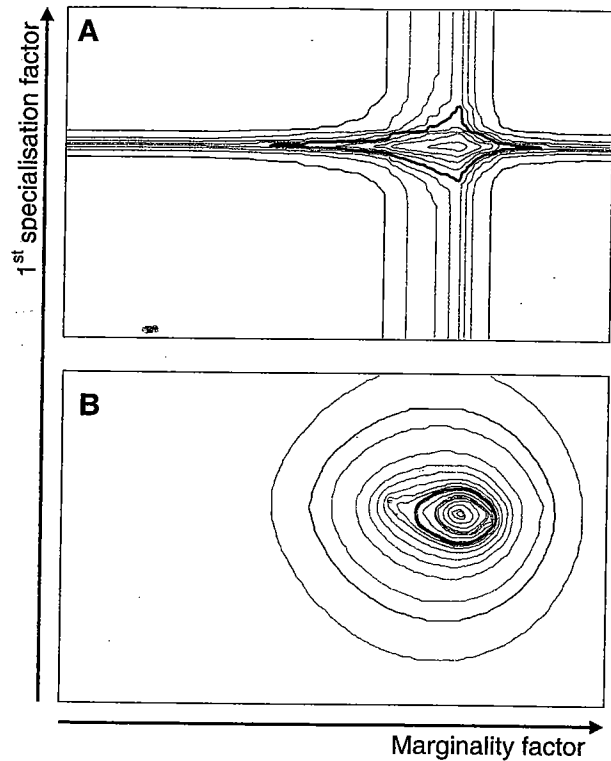

Figure 3. Habitat suitability field symbolised by its isolines in the same environmental space as Figure 2, as modeled by the median (A) and the geometric mean (B) algorithms. The plain isolines correspond to core and marginal envelopes. Only two dimensions are shown here although the actual modeling was conducted in a 7-dimensional space.

Discussion

A behavioral particularity of the bearded vulture made the modeling of its habitat a challenge. Its carrion-based diet and great flying capacity allow it to explore daily a very wide area (up to 37 km linear distance from nest site for adults) (Arlettaz and others 2002). It can therefore be observed flying over a great diversity of habitats, many of them not at all or poorly suitable. Of course, most of the observations will be located close to the best habitats, and any model must be based on this property and not be too outlier-sensitive. In this respect, as both algorithms are being based on observation density in the environmental space, they give equally good results. During preliminary phases of this work, we tested two other algorithms: the first one, inspired from the DOMAIN algorithm (Carpenter and others 1993), was based on the distance to the closest observation point in the environmental space. The second one was similar to the geometric mean algorithm but used the harmonic mean instead; the latter was considered before the geometric mean because of its
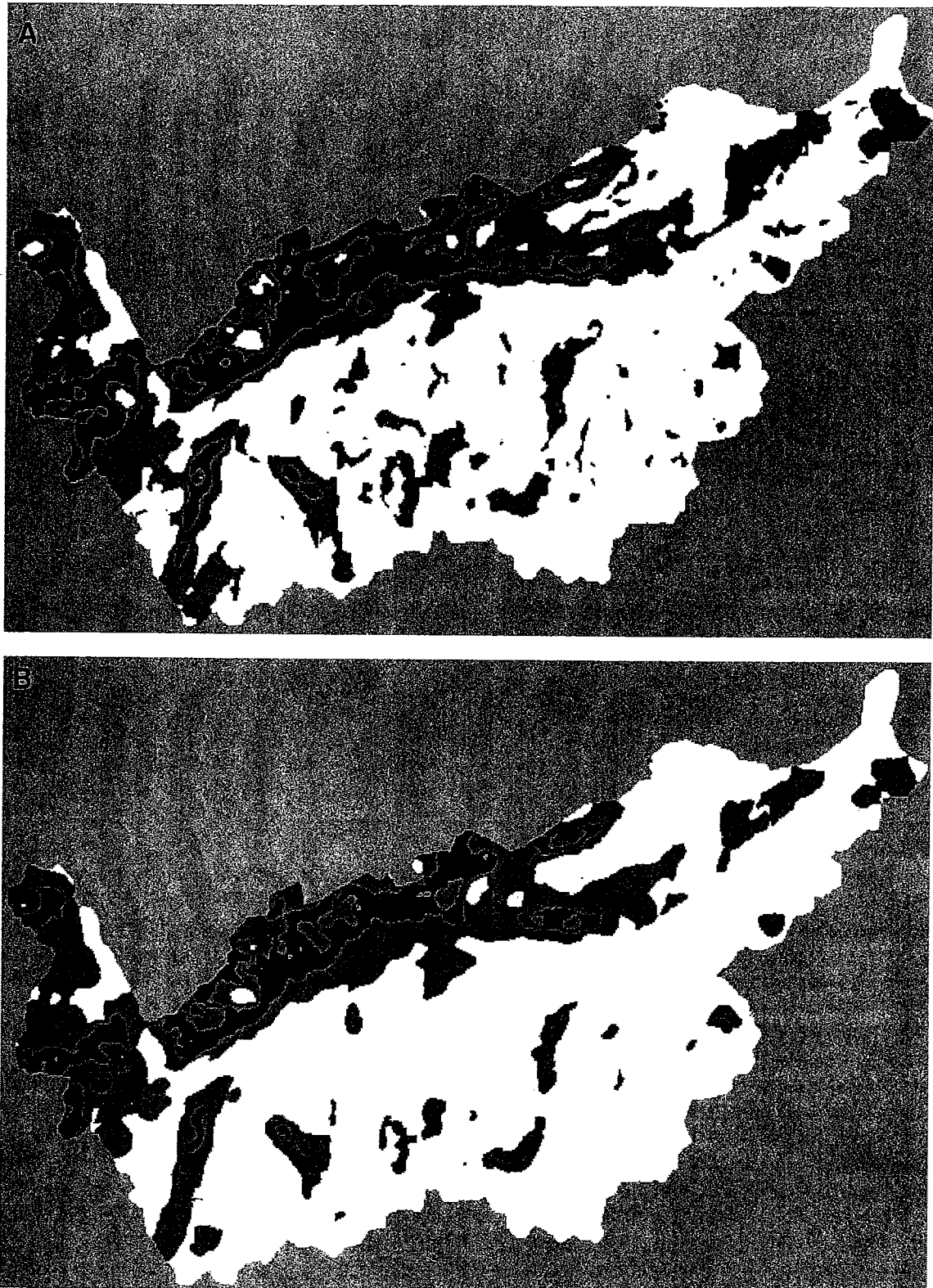
Figure 4. Habitat suitability maps resulting from the median (A) and geometric mean (B) algorithms. They show the spatial distribution of the core (black), marginal (dark gray), and unsuitable habitats (white).

well-known application to model species home ranges in the geographic space (Dixon and Chapman 1980, White and Garrott 1990). Both algorithms were shown to give too much importance to isolated (in the environmental space) observations and were therefore not able to provide a good generalization of the habitat. They would probably achieve better results on less vagrant species, as was shown by Carpenter and others (1993) on plants. We argue that, for a far-ranging animal like the bearded vulture, environmental space can be substituted for geographical space for modeling, providing that observation density rather than observation only is taken into account. Indeed, although vultures can be seen flying over a wide range of habitat, our data show that it nevertheless tends to stay much more frequently on particular parts of the available area that translate into well defined spots in the environmental space.

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 colonizing the mountains north of the Rhone. This shift in geographic space (Figure 1) is reflected in the environmental space by a bimodal pattern: the left-hand isolated optimum (Figure 2A) corresponds to the suboptimal valley firstly colonized and the twin main optima (Figure 2B) to the more recently selected area; the points scattered throughout the environmental space correspond to occasional observations of birds in unsuitable habitats. By conducting a separate analysis on each period, it was possible to draw clues about the ecological requirements and behavior of immature and subadult vultures (Hirzel and Arlettaz,submitted): immatures were mainly attracted by areas with high prey density (ibex and sheep) and stayed closer to the release site; the subadults were more driven by the need to look for areas favoring breeding and nestling (chiefly calcareous areas). However, no pair has yet nested in Switzerland and, therefore, the observed distribution is genuinely due to utilization preferences rather than merely to flight routes to and from the nest sites.

This bimodal distribution misleads the median algorithm, as it would for any model making too stringent assumptions on the environmental space landscape as it assumes that the factor distributions are unimodal and roughly symmetrical. Although this assumption is generally correct, there are cases where it is not (Austin 2002, Oksanen and Minchin 2002). Asymmetry in the environmental space may arise if the study area is situated in a marginal part of the species repartition where the most suitable environmental conditions are scarce and where the species has to settle down in suboptimal areas. This could be due to habitat fragmentation, as it has been shown to reduce niche breadth (Gehring and Swihart 2003, Swihart and others 2003), which probably also has an effect on its symmetry. Asymmetry may also result from interspecific competition or climate change (Austin 1992, 2002). Anyway, strong skewness will cause the median algorithm to overestimate the suboptimal areas and underestimate the most suitable ones. Multimodality is probably rarer but may be caused by: (1) historical events—a spreading species settling down in a suboptimal area before it discovers a better one, (2) interaction with a more competitive species that pushes it away from its intrinsic optimum (Austin 2002), or (3) by adaptation to two or more kinds of environment (Kawata 2002). In such cases, as the median is prone to lie somewhere between the optima, this algorithm produces incorrect results.

These two characteristics, namely sporadicity and multimodality, generate contradictory constraints: while multimodality needs a closely fitting model, outlier sporadicity must be smoothed by a more parametric one. This paper shows that a model based on the distance geometric mean is a good trade-off between the two ever-competing constraints of generality and precision.

According to these results, should the median algorithm nevertheless be abandoned? In many cases, observation distribution in the environmental space is unimodal and grossly symmetric around the optimum, and it has been applied successfully to many species (Reutter and others 2003). However, its main advantage is its simplicity and, therefore, its computing speed. With large sets of observations, distance-based algorithms such as the geometric mean are far slower than the median algorithm. While this is not a great problem when computing only one model, it can be time-consuming when Monte Carlo statistics are involved, such as the jack-knife cross-validation procedure applied here. A good practice would be to perform a preliminary analysis using the median algorithm and to look at the species distribution on the factors, particularly at the marginality factor as the potential multimodality or asymmetry patterns are most likely to show on this factor. Anytime the distribution is bimodal or, more generally, the median does not seem to be a good approximation of optimal habitat, the geometric mean algorithm should preferred.

Acknowledgements

the Swiss Ornithological Institute (Hans Schmid), and the Foundation Pro Bartgeier (Anne-Marie and Chasper Buchli) which gave us free access to the bearded vulture databases. Special thanks go to Christian Glenz and P. A Oggier. We are also grateful to Nicolas Perrin, Antoine Guisan, and William Hargrove for their interesting insights during the conception of these algorithms, and to Anthony Lehmann, Virginia H. Dale, and two anonymous reviewers for their keen comments about our manuscript.

References

Arlettaz, R., Posse, B. and Crettenand. Y. 2002. The bearded vulture in Valais in 1999. Pages 58–59 in Bearded vulture reintroduction into the Alps, Annual report 2000.

Austin, M. P. 1992. Modeling the environmental niche of plants—implications for plant community response to elevated CO_2 levels. *Australian Journal of Botany* 40:615–630.

Austin, M. P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling* 157:101–118.

Busby, J. R. 1988. Potential impacts of climate change on Australia's flora and fauna. Pages Pages 387–398 in G. I. Pearman Eds, 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. Eds, Nature conservation: cost effective biological 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.

Dixon, K. R., and J. A. Chapman. 1980. Harmonic mean measure of animal activity areas. *Ecology* 61:1040–1044.

Fielding, A. H., and J. F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24:38–49.

Gehring, T. M., and R. K. Swihart. 2003. Body size, niche breadth, and ecologically scaled responses to habitat fragmentation: mammalian predators in an agricultural landscape. *Biological Conservation* 109:283–295.

Gower, J. C. 1971. General coefficient of similarity and some of its properties. *Biometrics* 27:857.

Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135:147–186.

Hirzel, A. H., V. Helfer, and F. Métral. 2001. Assessing habitat-suitability models with a virtual species. *Ecological Modelling* 145:111–121.

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.

in Division of Conservation Biology, Bern. URL: http://www.unil.ch/biomapper.

Hutchinson, G. E. 1957. Concluding remarks. *Cold Spring Harbor Symposium on Quantitative Biology* 22:415–427.

Kawata, M. 2002. Invasion of vacant niches and subsequent sympatric speciation. *Proceedings of the Royal Society of London Series B-Biological Sciences* 269:55–63.

Manly, B. F. J. 1991. Randomization and Monte Carlo methods in biology. Chapman and Hall, London.

Oksanen, J., and P. R. Minchin. 2002. Continuum theory revisited: what shape are species responses along ecological gradients?. *Ecological Modelling* 157:119–129.

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.

Sokal, R. R., and F. J. Rohlf. 1981. Biometry: The principles and practice of statistics in biological research. 2. W. H. Freeman & Co., New York.

Swihart, R. K., T. M. Gehring, M. B. Kolozsvary, and T. E. Nupp. 2003. Responses of 'resistant' vertebrates to habitat loss and fragmentation: the importance of niche breadth and range boundaries. *Diversity and Distributions* 9:1–18.

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.

White, G. C., and R. A. Garrott. 1990. Analysis of wildlife radiotracking data. Academic Press, San Diego.

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. 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 losing 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 ecogeographical 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.

Each cell can thus be represented by a point in the multidimensional space of the variables. If distributions are multinormal, their distribution will have the shape of a hyperellipsoid. The cells where the focal species were observed constitute a subset of the global distribution and define a smaller hyperellipsoid 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 standardized distance between these centroids.

The two ellipsoids are then projected onto a hyperplane perpendicular to the marginality factor. This ensures that subsequent factors will be uncorrelated to the first and removes one dimension from the environmental space. The first of the specialization factors is obtained by computing the axis that maximizes the ratio of global variance over species variance. The following specialization factors are then extracted in turn by repeating this process, each step removing one dimension from the space, until all V factors are extracted. The amount of specialization accounted for by a factor is necessarily lower than that due to the one previously extracted. Furthermore, all specialization factors are 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.

The mathematical demonstration of this procedure is developed by Hirzel and others (2002).