Ecology, 89(2), 2008, pp. 555–566

THE FACTORIAL DECOMPOSITION OF THE MAHALANOBIS DISTANCES IN HABITAT SELECTION STUDIES

C. Calenge,[1,2,3] G. Darmon,[1] M. Basille,[1] A. Loison,[1] and J.-M. Jullien[2]

[1]*Laboratoire de Biométrie et Biologie Evolutive (UMR 5558); CNRS; Université Lyon 1, 43 Bd 11 Nov, 69622, Villeurbanne Cedex, France*
[2]*Office National de la Chasse et de la Faune Sauvage, 95 Rue Pierre Flourens, 34000 Montpellier, France*

Abstract. The Mahalanobis distances have been introduced in habitat selection studies for the estimation of environmental suitability maps (ESMs). The pixels of raster maps of a given area correspond to points in the multidimensional space defined by the mapped environmental variables (ecological space). The Mahalanobis distances measure the distances in this space between these points and the mean of the ecological niche (i.e., the hypothesized optimum for the species) regarding the structure of the niche. The map of these distances over the area of interest is an estimated ESM. Several authors recently noted that the use of a single optimum for the niche of a species may lead to biased predictions of animal occurrence. They proposed to use instead a minimum set of basic habitat requirements, found by partitioning the Mahalanobis distances into a restricted set of biologically meaningful axes. However, the statistical approach they proposed does not take into account the environmental conditions on the area where the niche was sampled (i.e., the environmental availability), and we show that including this availability is necessary. We used their approach as a basis to develop a new exploratory tool, the Mahalanobis distance factor analysis (MADIFA), which performs an additive partitioning of the Mahalanobis distances taking into account this availability. The basic habitat requirements of a species can be derived from the axes of the MADIFA. This method can also be used to compute ESMs using only this small number of basic requirements, therefore including only the biologically relevant information. We also prove that the MADIFA is complementary to the commonly used ecological-niche factor analysis (ENFA). We used the MADIFA method to analyze the niche of the chamois *Rupicapra rupicapra* in a mountainous area. This method adds to the existing set of tools for the description of the niche.

Key words: chamois; ecological-niche factor analysis; environmental suitability maps; exploration; French Alps; habitat selection; Mahalanobis distances factor analysis; niche; Rupicapra rupicapra.

Introduction

The detailed knowledge of species distribution is of major concern for a large range of ecological topics. Among the tools available to improve this knowledge, environmental suitability maps (ESMs) occupy the first place (Guisan and Zimmermann 2000, Manly et al. 2002, Elith et al. 2006). Such maps are essential for decision making in wildlife management (Knick and Rotenberry 1998) and for building conservation plans (Araujo and Williams 2000).

Most methods developed to build ESMs rely on the concept of ecological niche (Guisan and Zimmermann 2000). These maps are generally estimated using a sample of species occurrences on an area mapped for several environmental variables (e.g., elevation, slope, vegetation). Each environmental variable defines a dimension of a multidimensional space, hereafter termed "ecological space." The values of these variables can be determined for each species occurrence, so that the whole set of occurrences defines a cloud of points in the ecological space, the species niche. Environmental suitability mapping implies the computation of one environmental suitability index for each pixel of the map, based on the position of the corresponding point in the ecological space relative to the species niche. These indices are then mapped in the geographical space to provide an ESM.

The commonly used Mahalanobis distance between the available point and the mean of the niche is such an index (Mahalanobis 1948, Clark et al. 1993, Knick and Dyer 1997, Knick and Rotenberry 1998, Corsi et al. 1999, Farber and Kadmon 2003, Cayuela 2004, Thompson et al. 2006). The mean of the niche is supposed to reflect the environmental conditions optimal for the studied species. The Mahalanobis distance for a given point expresses the distance between this point and the species optimum in the ecological space, regarding the niche structure (see Appendix A for a precise graphical description of these distances). If we assume that smaller distances correspond to areas that are more likely to be occupied by the species, the Mahalanobis distances can

Manuscript received 19 October 2006; revised 6 June 2007; accepted 22 June 2007. Corresponding Editor: N. G. Yoccoz

[3] E-mail: calenge@biomserv.univ-lyon1.fr

be mapped over the study area to provide a reliable ESM.

Recently, several authors noted that the mean of the niche of a species on a given study area can be a poor proxy for its optimum (Dunn and Duncan 2000, Rotenberry et al. 2002, 2006, Browning et al. 2005). More suitable characteristics of the environment found in another area, but not in the original one, will be characterized by large Mahalanobis distances, and therefore low estimated suitability. The Mahalanobis distances may therefore lead to biased predictions of animal occurrence under different environmental conditions. These authors proposed to use, instead of this optimum, a minimum set of basic habitat requirements. They advocated that the variables that maintain a consistent value where the species occur (i.e., the variables with a low "used" variance) are those most likely to be associated with basic habitat requirements. For this reason, they argued that the last axes of a principal component analysis (PCA) of the niche, on which the variance is the smallest, can be used to define this basic set. Moreover, they demonstrated that this PCA is a natural way to partition the Mahalanobis distances. Therefore, these authors recommended estimating ESMs by computing a reduced-rank Mahalanobis distance for each pixel of the map of the study area, by considering only this restricted set of principal components. They consider this statistic as the distance from the pixel to this minimum set of basic requirements.

However, although this linear partitioning of the Mahalanobis distance relies on both solid mathematical bases and sound biological issues, it is also problematic. The PCA recommended by these authors is performed on the table giving the value of the environmental variables (columns) in the sites used by the species (rows), without consideration of the availability of the environmental variables. Note that this table is standardized before the PCA is applied, so that all the environmental variables have a unit variance. This preliminary operation is necessary, as the variables may not be measured on the same scale (e.g., the elevation measured in meters and slope measured in percent). However, this scaling has an unexpected consequence: maximizing the variance of the standardized niche on the first axes of the PCA is just a way of maximizing the sum of the squared correlations between the environmental variables and the first axis (Legendre and Legendre 1998).

However, the fact that some environmental variables are strongly correlated among each other does not imply that these variables cannot be used to define a basic set of required habitats. For example, hydrobiologists often measure the velocity, the depth, and the flow of a stream when they want to study the niche of a fish species (e.g., Mäki-Petäys et al. 1997). These variables are often strongly correlated among each other, even when the correlations are computed only with the sites used by the species. These variables are therefore likely to define the first axis of the PCA of the niche. However, they are strong limiting factors for many species, in the sense that the range of variation actually experienced by the species is very small relative to the range that could be potentially encountered by the species.

The crucial point here is that the identification of variables with a "low" variance implies that we know what a "normal" variance is for these variables: a reference value is needed. Actually, the used sites are generally sampled on a given area, which defines the context in which the niche takes place. The whole set of pixels of this area defines a cloud of "available points" in the ecological space, of which the niche is a subset. The shape of the niche in the ecological space is partly the result of the influence of this context. Actually, we defend the idea that the identification of the required habitat for a species distribution from a sample of used sites should also take into account the environmental availability at the time of sampling in some way.

However, the biological issue raised by Rotenberry et al. (2002, 2006) is important. The definition of a restricted set of basic habitat requirements could improve the predictive capabilities of ESMs based on the Mahalanobis distances. In this paper, we used the work of Rotenberry et al. (2002, 2006) as a basis to solve the problem of the identification of the basic habitat requirements. We therefore developed a new exploratory approach to tackle the problem, which we called the "Mahalanobis Distances Factor Analysis" (MADIFA). This approach also performs an additive partitioning of the Mahalanobis distances, but the first components of the analysis now explain most of the Mahalanobis distances for the set of available points on a given area. The factorial maps of these axes allow both the exploration of the niche in the ecological space and the identification of the environmental variables corresponding to basic habitat requirements. The factorial axes can also be used to compute ESMs on a lower number of dimensions (and therefore with increased generality) that take into account a large part of the niche restriction. We illustrate how this analysis may find its place among other exploratory tools of the niche with the analysis of the niche of the chamois (*Rupicapra rupicapra*) in a mountainous environment.

The Computation of the Mahalanobis Distances

We assume that the values of P environmental variables are known for N pixels (where N can be a random sample or the whole set of pixels of a map). We consider here that the N available pixels have the same weight in the analysis, contained in the $N \times N$ (rows × columns) diagonal matrix $\mathbf{D} = \mathrm{Diag}(1/N)$. Moreover, we consider a set of N utilization weights, summing to one, which reflects the use of the N pixels by the focus species. For example, these weights may correspond to the proportion of locations of the studied species in the pixels of the map. These weights are stored in an $N \times N$

diagonal matrix $\mathbf{D}_p$. In the rest of this paper, we will term "available pixels" the whole set of N pixels of the analysis, and "used pixels" or "niche" the set of pixels for which the utilization weights are greater than zero.

Let the matrix $\mathbf{Z}$ contain the value of the P environmental variables (columns) in each one of the N available pixels (rows). The matrix $\mathbf{Z}$ is centered and scaled for the weighting $\mathbf{D}_p$ (i.e., respectively, the origin of the space defined by the columns of $\mathbf{Z}$ is located at the mean and the variance is 1 for all columns of $\mathbf{Z}$). Finally, let $\mathbf{\Sigma} = \mathbf{Z}^\mathsf{T}\, \mathbf{D}_p\, \mathbf{Z}$ be the correlation matrix as the columns of $\mathbf{Z}$ have a unit variance (where $\mathbf{Z}^\mathsf{T}$ is Tthe transpose of $\mathbf{Z}$).

The squared Mahalanobis distance D_i^2 between any available point i (associated to a pixel in the geographical space) and the mean of the niche provides an index of the environmental suitability at this place. Let $\mathbf{Z}_{i\cdot}$ be the row vector containing the values of the P environmental variables for the ith pixel (that is, the ith row of the matrix $\mathbf{Z}$). In these conditions, the squared Mahalanobis distance between the point i and the mean of the niche can be computed with

$$D_i^2 = \mathbf{Z}_{i\cdot}\mathbf{\Sigma}^{-1}\mathbf{Z}_{i\cdot}^\mathsf{T}. \tag{1}$$

Linear Partitioning of the Mahalanobis Distances: The Point of View of Rotenberry et al. (2002, 2006)

Rotenberry et al. (2002, 2006) noted that the computation of these distances relies on the computation of the inverse of the matrix $\mathbf{\Sigma}$ (Eq. 1). This computation may be performed by its diagonalization (i.e., the computation of its eigenvectors and eigenvalues). More formally,

$$\mathbf{\Sigma} = \mathbf{A}\mathbf{\Lambda}\mathbf{A}^\mathsf{T}$$

where the matrix $\mathbf{\Lambda}$ is the diagonal matrix containing the P eigenvalues λ_j of the matrix $\mathbf{\Sigma}$, i.e., Diag($\lambda_1, \lambda_2, \ldots, \lambda_p$), and $\mathbf{A}$ is the matrix containing the P eigenvectors $\boldsymbol{\alpha}_j$ of the matrix $\mathbf{\Sigma}$ concatenated by columns, i.e., [$\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_P$]. The inverse of the matrix $\mathbf{\Sigma}$ is given by the following (Harville 1997):

$$\mathbf{\Sigma}^{-1} = \mathbf{A}\Lambda^{-1}\mathbf{A}^\mathsf{T}.$$

Consequently, the Mahalanobis distance between the point i and the mean of the niche can be computed using

$$D_i^2 = \mathbf{Z}_{i\cdot}\mathbf{A}\Lambda^{-1}\mathbf{A}^\mathsf{T}\mathbf{Z}_{i\cdot}^\mathsf{T}. \tag{2}$$

Rotenberry et al. (2002, 2006) noted that this formula provides a natural way of partitioning the Mahalanobis distances, as it is related to the principal components analysis (PCA) of the niche (i.e., a PCA of the table $\mathbf{Z}$ using the matrix $\mathbf{D}_p$ as row weights; as in Fig. 1B). The axes of this PCA correspond to the eigenvectors of $\mathbf{\Sigma}$ (i.e., $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$, etc.). The first axes represent the directions in the ecological space for which the niche width is maximal. The variance of the niche projected onto a given axis j of this PCA is the corresponding eigenvalue λ_j. Note that because $\mathbf{Z}$ has been scaled, this maximized variance is just the sum of squared correlations between the environmental variables and the axis j of the PCA (Legendre and Legendre 1998). The vector $\mathbf{Z}_{i\cdot}$ contains the coordinates of the available point i in the ecological space. Therefore the coordinate of the available point i projected onto the jth axis of the PCA is computed by $\mathbf{Z}_{i\cdot}\boldsymbol{\alpha}_j$. The normed coordinate b_{ij} of the point i on the jth factorial axis corresponds to the raw coordinate divided by the standard deviation of the niche on this axis. Then, using Eq. 2, it is straightforward to show that the Mahalanobis distances can be computed by the sum of the squared b_{ij}:

$$D_i^2 = \sum_{j=1}^{P} b_{ij}^2 = \sum_{j=1}^{P}\left(\frac{\mathbf{Z}_{i\cdot}\boldsymbol{\alpha}_j}{\sqrt{\lambda_j}}\right)^2. \tag{3}$$

Rotenberry et al. (2002, 2006) advocated the use of a limited set of PCA axes to compute reduced-rank Mahalanobis distances. They noted that the first axes of the PCA are unlikely to describe required habitats, precisely because they thought that the large variance on these axes indicated that the ecological variation experienced by the species was large (whereas this variance is just the sum of squared correlation with the environmental variables). They proposed instead to compute the reduced-rank Mahalanobis distances using the last eigenvectors of the PCA, arguing that the dimensions on which the niche is the narrowest are likely to describe required habitats. For example, using the last R axes of the PCA, the reduced-rank squared Mahalanobis distances $\tilde{D}_i^2$ is computed using

$$\tilde{D}_i^2 = \sum_{j=P-R}^{P} b_{ij}^2.$$

Some Refinements of this Point of View: The MADIFA

The three steps to perform the MADIFA

We develop here a new partitioning of the Mahalanobis distances, which identifies the directions in the ecological space for which the niche is the narrowest in comparison to the width of the cloud of available points (see Fig. 1). We call it the "Mahalanobis Distances Factor Analysis" (MADIFA). This analysis is performed in three steps. The first two steps of this analysis are exactly the approach proposed by Rotenberry et al. (2002, 2006).

A PCA is first performed on the table $\mathbf{Z}$ using the matrix $\mathbf{D}_p$ as row weights, which returns the directions partitioning the variance of the standardized niche into orthogonal components (Fig. 1B), i.e., the set of eigenvectors $\boldsymbol{\alpha}_j$ and of eigenvalues λ_j ($j = 1, \ldots, P$) of the matrix $\mathbf{\Sigma}$ as defined in Eq. 2. Second, the ecological space is distorted: the correlation structure is removed by rescaling the variance of all axes to one (Fig. 1C). The scores of the available pixels in this distorted space are

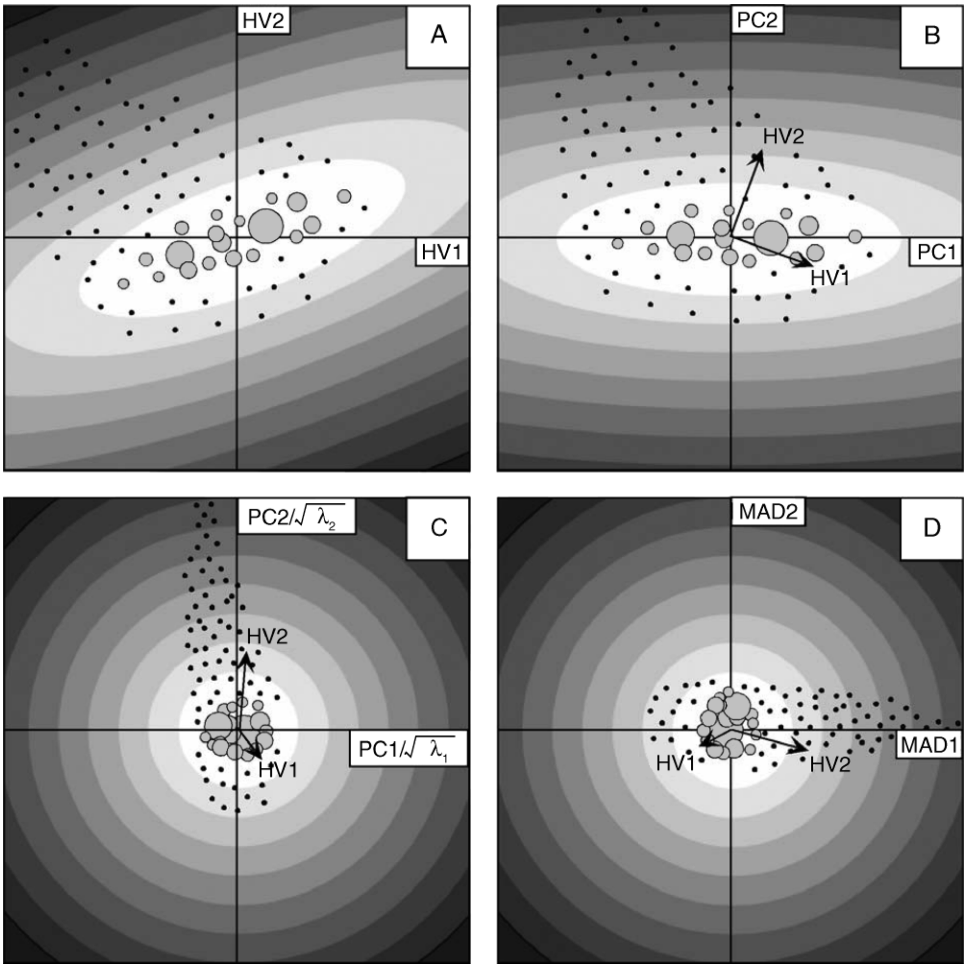

FIG. 1. The MADIFA procedure (see *The three steps to perform the MADIFA*). The black dots are points considered as available to the species. To each available point is associated one utilization weight proportional to its use by the species (indicated by a gray circle with an area proportional to this weight). The values of the Mahalanobis distance to the optimum of the niche are indicated by gray levels (i.e., the darker the shade, the farther from optimum). (A) The ecological space is defined by two environmental variables, HV1 and HV2, and is centered on the mean of the niche. (B) The first principal component analysis (PCA) of the niche (PC1 and PC2 are the principal components). (C) The scores of the points on the two principal components are divided by the square root of their respective eigenvalues. (D) The second PCA (not centered) maximizes the mean-squared Mahalanobis distances between the available points and the mean of the niche on the first axes, MAD1 and MAD2.

stored in the $N \times P$ matrix $\mathbf{B}$:

$$\mathbf{B} = \mathbf{Z}\mathbf{A}\mathbf{\Lambda}^{-1/2}. \tag{4}$$

The matrix $\mathbf{B}$ contains the normed scores b_{ij} as defined in Eq. 3 (Rotenberry et al. 2002, 2006).

Thereafter, we add another step to this approach: we perform a PCA on matrix $\mathbf{B}$ using the uniform row weights stored in $\mathbf{D}$. This second PCA is the core of the MADIFA, and we show hereafter that it returns linear combinations of the environmental variables so that the width of the niche is the smallest in comparison to the width of the cloud of available points (Fig. 1D).

Mathematical properties of the second PCA

The matrix being diagonalized is $\mathbf{G} = \mathbf{B}^{\mathrm{T}}\ \mathbf{D}\ \mathbf{B}$. This analysis returns a set of P orthogonal eigenvectors $\mathbf{v}_k$ stored in a matrix $\mathbf{V}$, and P corresponding eigenvalues θ_k stored on the diagonal of the matrix $\mathbf{\Theta}$, so that $\mathbf{G} = \mathbf{V}\mathbf{\Theta}\mathbf{V}^{\mathrm{T}}$. The pixel scores are computed by $\mathbf{L} = \mathbf{B}\mathbf{V}$:

$$\mathbf{L} = \mathbf{Z}\mathbf{A}\mathbf{\Lambda}^{-1/2}\mathbf{V}. \tag{5}$$

This formula summarizes the three steps of the MADIFA (Fig. 1): the factorial axes of this analysis are found after a rotation (matrix $\mathbf{A}$), a distortion (matrix $\mathbf{\Lambda}^{-1/2}$), and another rotation (matrix $\mathbf{V}$) of the cloud of available points in the ecological space (matrix $\mathbf{Z}$). All these transformations of $\mathbf{Z}$ can be summarized in a matrix $\mathbf{C} = \mathbf{A}\mathbf{\Lambda}^{-1/2}\mathbf{V}$. The pixels scores are the linear combinations of the environmental variables (i.e., $\mathbf{L} = \mathbf{Z}\mathbf{C}$).

The value maximized on the first axes of the MADIFA is equal to the following:

$$\theta_j = \frac{1}{N}\sum_{i=1}^{N} l_{ij}^2 = \frac{\sum_{i=1}^{N}\frac{1}{N}(l_{ij} - \bar{l}_j^u)^2}{\sum_{i=1}^{N} u_i(l_{ij} - \bar{l}_j^u)^2} \qquad (6)$$

where l_{ij} is the score of the pixel i on the jth axis of the MADIFA, $\bar{l}_j^u$ is the mean of the scores of the used pixels on the jth axis of the analysis, and u_i is the utilization weight associated with the pixel i. This result derives from the observation that the used variance (denominator of θ_j) is equal to 1 on the axes of the MADIFA, and that $\bar{l}_j^u = 0$ (as the origin of the ecological space is the mean of the niche).

Thus, the denominator of θ_j is the variance of the niche on the first axis of the MADIFA. However, the numerator is not a variance: it is the mean of the squared deviations of the available points from the mean of the scores of used points. Consequently, the MADIFA indicates the directions where the niche is the narrowest (low variance) compared to the width of the distribution of available points. This direction is likely to define a basic habitat requirement.

We show in Appendix B that

$$D_i^2 = \sum_{j=1}^{P} l_{ij}^2. \qquad (7)$$

Note that this result implies that the sum of the eigenvalues θ_j over all the axes j of the analysis is equal to the mean of the squared Mahalanobis distances for the available pixels. It is therefore possible to compute the proportion of the mean-squared Mahalanobis distances explained by each axis.

Now, like Rotenberry et al. (2002, 2006), we can compute reduced-rank squared Mahalanobis distances with the set of R first axes (chosen so that the variance of the niche is the smallest as compared to the variance of the available points), reflecting the distance between the available points and the set of basic habitat requirements. From Eq. 7, one can derive the reduced-rank squared Mahalanobis distance:

$$\tilde{D}_i^2 = \sum_{j=1}^{R} l_{ij}^2. \qquad (8)$$

The scores of the pixels on the axes of the MADIFA can be used to draw factorial maps to identify the structures of the niche in the ecological space (as in Fig. 1D). Alternatively these scores can be used to map reduced-rank Mahalanobis distances over the area, to provide clearer and sharpened environmental suitability maps (ESMs; using Eq. 8). The biological meaning of the factorial axes can be found either by using the coefficients in **C** or the correlations with the original environmental variables.

The MADIFA is programmed in the function "madifa()" of the free package adehabitat (Calenge 2006) for the R software (R Development Core Team 2005). It can be used as a classical exploratory tool (Legendre and Legendre 1998) to draw a conceptual model of the studied biological system.

The MADIFA and the ecological-niche factor analysis

The MADIFA is closely related to the ecological-niche factor analysis (ENFA) developed by Hirzel et al. (2002). Indeed, these authors noted that basic habitat requirements are likely to be associated with the directions of the ecological space where the variance of the niche is very small in comparison to the variance of the available points. The ratio of these two variances computed for a given variable is an index of the specialization of the species on this variable. The ENFA is a factor analysis of the niche maximizing this index on the first axis. More formally, for a given axis j, the specialization ratio S is equal to

$$S(w_j) = \frac{\sum_{i=1}^{N}\frac{1}{N}(w_{ij} - \overline{w}_j^a)^2}{\sum_{i=1}^{N} u_i(w_{ij} - \overline{w}_j^u)^2} \qquad (9)$$

where w_{ij} is the score of the ith pixel on the jth axis of the ENFA, $\overline{w}_j^a$ is the mean of the scores of available points on the jth axis of the ENFA, and $\overline{w}_j^u$ is the mean of the scores of the used points on the same axis. Note that $S(w_j)$ is very similar to θ_j (compare Eq. 6 and Eq. 9). The only difference is that the former uses the variance of available points as a measure of the width of the distribution of available points, while the latter uses the mean of the squared deviation of available points from the mean of the scores of used points.

Maximizing the ratio $S(w_j)$ is possible only if the marginality vector has first been extracted from the data (i.e., the vector connecting the mean of the cloud of available points to the mean of the cloud of used points; Hirzel et al. 2002). However, the marginality vector is often biologically important, and several authors stressed the need to take into account this vector in the interpretation of the results (e.g., Hirzel et al. 2002). Consequently, the available and used points are projected onto this vector to define a marginality axis as a first step. The interpretation of the results of the ENFA includes the interpretation of the scores of used and available points on this marginality axis.

Note that the ratio θ_j maximized by the axes of the MADIFA can be rewritten:

$$\theta_j = \frac{m_j^2}{v_j^2} + S(w_j)$$

where m_j^2 is the squared difference between the mean of the scores of used points and the mean of the scores of available points on the jth axis of the analysis (i.e., the marginality), and v_j^2 is the variance of the niche on the jth axis of the analysis. The MADIFA therefore

TABLE 1. Variables included in the "Mahalanobis Distances Factor Analysis" (MADIFA).

Abbreviation	Variable name
Elev	elevation
D.Alder	distance to alder woods
D.Screes	distance to screes
D.Forest	distance to forested areas
D.Fodder	distance to fodders
D.Brachy	distance to meadows dominated by *Brachipodium*
D.CarexF	distance to meadows dominated by *Carex ferruginea*
D.TallHe	distance to meadows dominated by tall herbs
D.Nardus	distance to meadows dominated by *Nardus* ssp.
D.SeCarS	distance to meadows dominated by *Sesleria* and *Carex sempervirens*
D.Rhodo	distance to moors dominated by *Rhododendron*
D.Trail	distance to recreational trails
Hydro	hydrography
Slope	slope
Sunshine	sunshine
Visib	visibility (area seen from each pixel, computed using Elev)
Visib1000	visibility computed within a radius of 1000 m

combines the marginality and the specialization into one single measure of niche restriction.

Thus, the ENFA may be used to complement the results of the MADIFA as it allows identification of the part of the Mahalanobis distances corresponding to the specialization and to the marginality, respectively. Used jointly, these two approaches lead to a more precise conceptual model elaborated for the niche of the focus species. The ENFA can also be used to draw factorial maps of the niche (Basille et al. 2008).

On the other hand, as the marginality axis does not have the same mathematical status as the specialization axes of the ENFA (the marginality axis is orthogonal to the specialization axes, but the specialization axes are not orthogonal among each other; Hirzel et al. 2002), it is often difficult to combine all these axes into one single index of environmental suitability. So far, existing methods trying to combine the marginality and specialization axes use ad hoc algorithms (Hirzel et al. 2002, Hirzel and Arlettaz 2003). Although these ENFA-based methods have proven to return biologically consistent environmental suitability maps (ESMs; e.g., Bryan and Metaxas 2007), the MADIFA is probably a better way to build environmental suitability maps: it returns axes, all with the same mathematical status, which can be combined into ESMs in a consistent manner.

APPLICATION: EXPLORATION OF HABITAT SELECTION BY THE CHAMOIS

We explored the habitat component of the niche of the chamois (*Rupicapra rupicapra*; see Plate 1) in open areas of the wildlife reserve of Les Bauges (French Alps, 45°25′ N, 6°5′ E; Fig. 2A). The data were collected during censuses carried out every year from 1994 to 2004 in June using the same protocol (flash counts; see e.g., Houssin et al. 1994). Volunteers and professionals working in various French wildlife and forest management organizations walked along 24 transects and looked around two fixed points, which were distributed over the reserve so that all open areas (i.e., nonforested areas) were visible to the observers. All transects were traveled simultaneously at dawn by teams of two observers, and each detected chamois group was located on a map of the reserve (precision of ~10 m). At the end of the census, hours and locations of observations were compared in order to delete the double counts. Because the study of habitat selection requires a homogeneous sampling effort, we used the upper elevation limit of the forests to delimit our study area (6430 ha dominated by open meadows located at an elevation >1200 m). Preliminary analysis showed that the number and the spatial distribution of the detected groups did not vary greatly among years (C. Calenge and G. Darmon, *unpublished data*). We therefore considered the pooled data set here to reduce these sampling fluctuations. During the seven years of the study, 650 chamois groups were detected (Fig. 2B). We split the data set in two, one for calibration (from 1994 to 2000; 400 groups detected), and one for validation (from 2001 to 2004; 250 groups detected). Seventeen environmental variables were included in the analysis of the chamois habitat (Table 1, Fig. 2C). These variables were supposed to reflect the chamois distribution, either because they reflect the location of secure areas (e.g., distance to trails, visibility, slope; von Elsner-Schack 1985), or because they represent vegetal associations in which the chamois may search for food (Ferrari et al. 1988, Garcia-Gonzalez and Cuartas 1996). Note that although we focused only on the chamois distribution in the open areas, we also included in the analysis the distance to forested areas, because these surrounding habitats may also influence the habitat use by the chamois in open areas (Hamr 1985).

We first investigated habitat selection using the calibration data set. Before the application of the MADIFA, we explored the structure of the environmental composition over the study area, using a principal component analysis of the table giving the values of the environmental variables (columns) in the pixels of the maps of the area (rows). One main pattern is highlighted (see Appendix C): the elevation, which is the variable best correlated with the first axis, affects the value of several environmental variables. Such an altitudinal structure was expected in this mountainous area. Areas close to the screes, to the meadows dominated by *Sesleria* and *Carex sempervirens*, and to the meadows dominated by *Carex ferruginea* are generally found at high elevations (Rameau et al. 2001).

We also performed a PCA restricted to the pixels where chamois were located (i.e., on its habitat). The altitudinal structure highlighted on the study area was

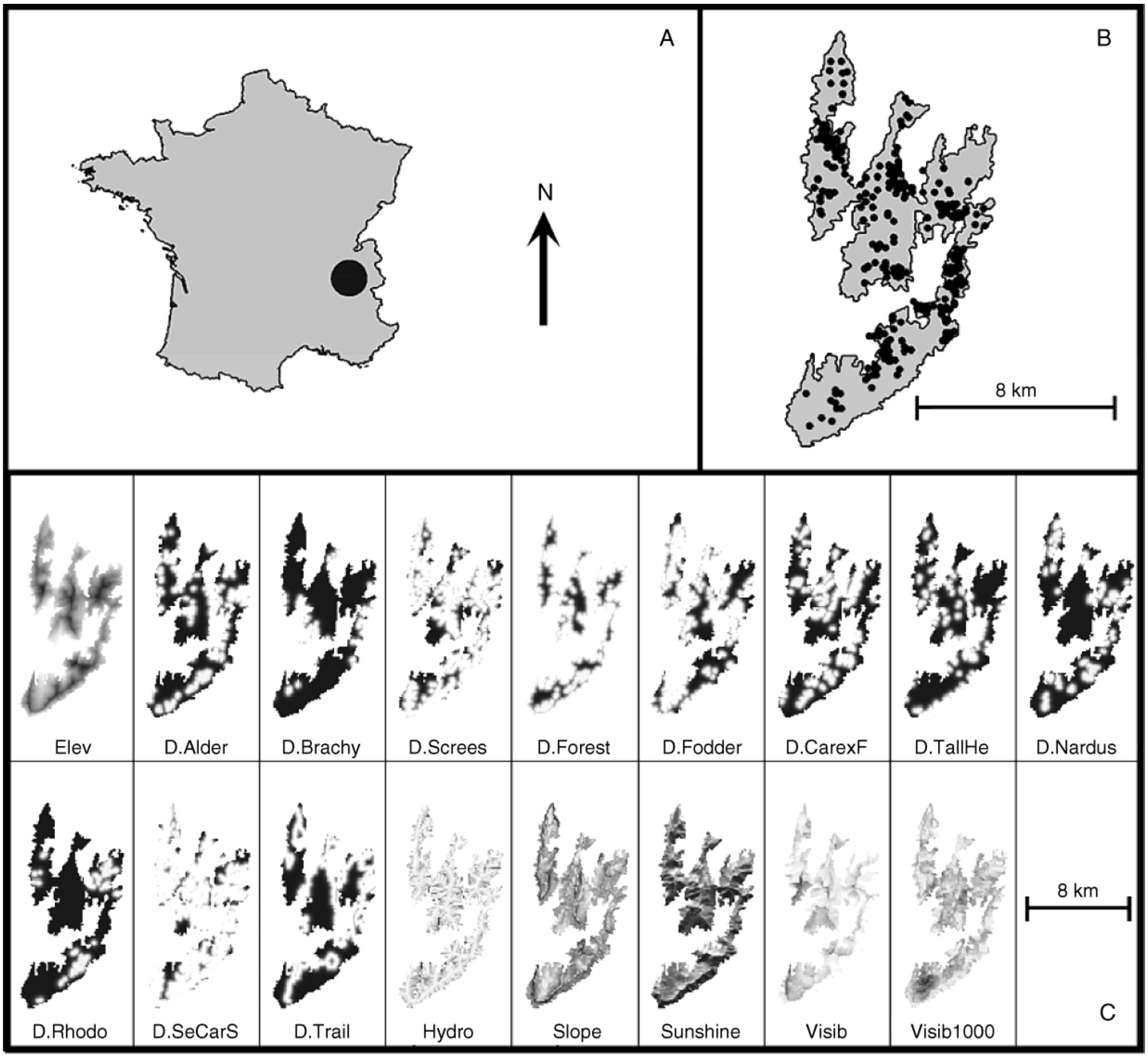

FIG. 2. (A) Location of the wildlife reserve of Les Bauges in France; (B) distribution of the chamois detected on the area from 1994 to 2000; and (C) maps of the 17 environmental variables over the area, where levels of each environmental variable increase from light to dark gray (see Table 1 for the full names of the variables).

also the main structure of the chamois habitat (Appendix C). The correlation between the first axis of the PCA of the available points and the first axis of the PCA of the habitat is very strong ($R = -0.87$). Actually, the altitudinal structure is so strong in the study area that it also affects the shape of the cloud of used points in the ecological space. However, the fact that the variance of used pixels is maximal on this direction does not imply that it does not describe a habitat required by the chamois, as shown next.

We then studied habitat selection of the chamois with the MADIFA. We first performed a preliminary Monte Carlo test to determine whether the habitat selection is significant in at least one direction of the ecological space. At each step of the process, we simulated a random habitat use by the chamois by generating a uniform distribution of 400 points over the study area, and we computed the first eigenvalue of the MADIFA of this simulated data set. We repeated this simulation 500 times to derive a distribution of eigenvalues under the hypothesis of random habitat use. We finally compared the first eigenvalue of the MADIFA of the observed 400 chamois groups to this simulated distribution to derive a P value. There is actually a highly significant habitat selection value ($\theta_1 = 3.7$, $P < 0.002$).

The proportion of the mean of the squared Mahalanobis distances explained by each axis j is measured by the corresponding eigenvalue θ_j. The exploration of these eigenvalues helps in choosing a number of axes to interpret (Fig. 3A). The MADIFA returned one main eigenvalue (15% of the mean of the squared Mahalanobis distances are explained on the first axis). The

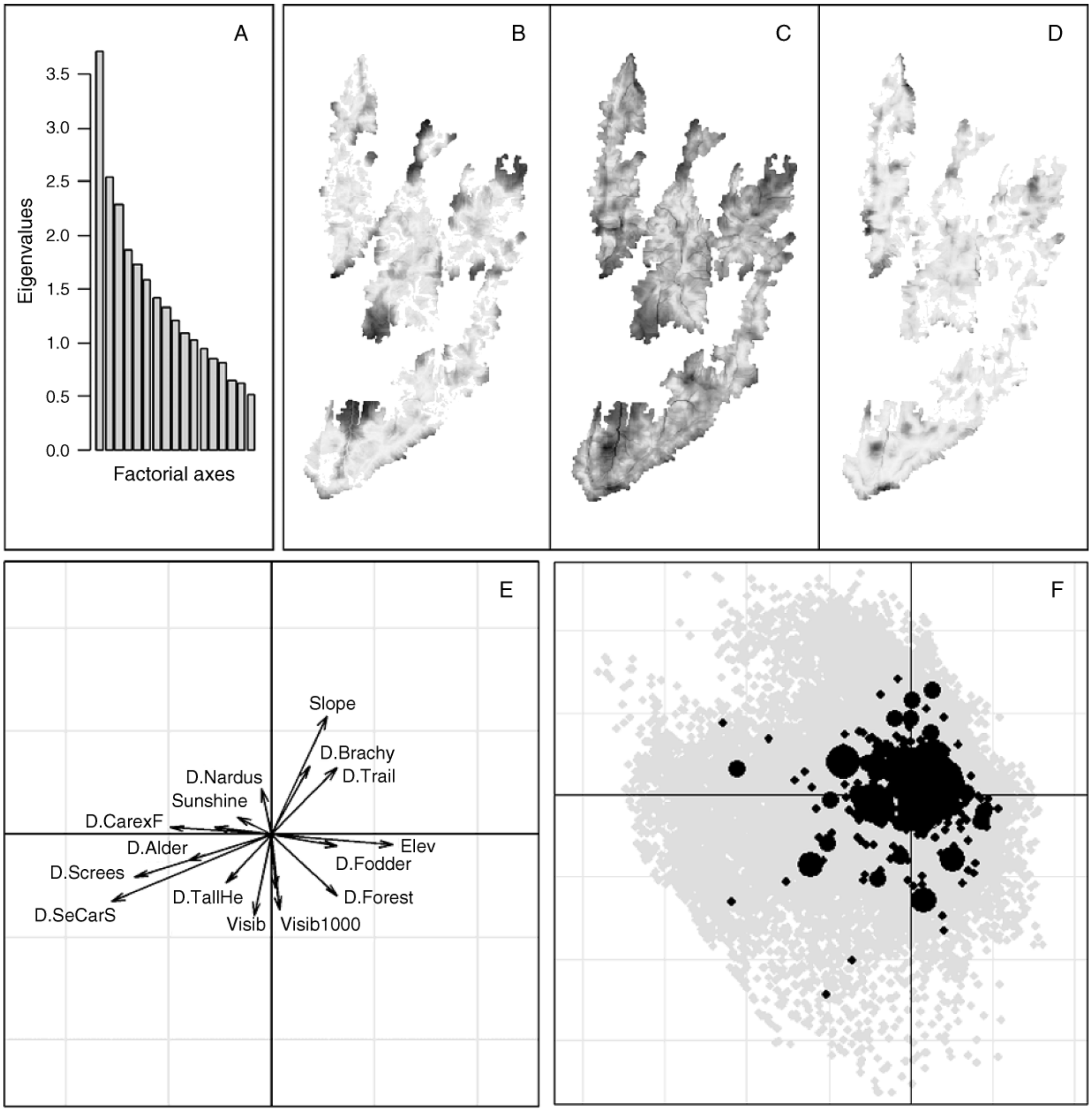

FIG. 3. Results of the MADIFA performed to analyze the chamois distribution with respect to the 17 environmental variables. Despite the fact that only one axis is highlighted by the analysis, results are presented for the first two axes in panels (E) and (F). For panels (B)–(D), levels of each environmental variable increase from light to dark gray. (A) Bar plot of the eigenvalues; (B) environmental suitability map of the area computed using the first axis of the MADIFA; (C) environmental suitability map of the area computed using the 17 environmental variables (full-rank Mahalanobis distances); (D) environmental suitability map of the area computed using the last seven axes of the PCA of the niche (method of Rotenberry et al. [2002, 2006]); (E) graph of the correlations between the environmental variables and the first (x-axis) and second (y-axis) axes of the MADIFA (see Table 1); and (F) factorial map of the ecological niche of the chamois on the first (x-axis) and second (y-axis) axes of the MADIFA. The gray points correspond to the available points (pixels of the maps), and their intensity of use is proportional to the area of the black points. The whole set of black circles defines the niche of the species.

percentage of the mean of the squared Mahalanobis distances explained by the following axes is much lower (10.5%, 9.5%, and 7.6% for the second, third, and fourth axis, respectively). We therefore focused our interpretation on the first axis of the MADIFA.

The biological meaning of this axis can be deduced from the correlation coefficients between the first axis of the MADIFA and the environmental variables (Fig. 3E). The positive scores on this axis correspond to areas located at high elevations (correlation between elevation and the first axis: $R = 0.59$), close to the screes (D.Screes, $R = -0.67$), and, above all, close to the meadows dominated by *Sesleria* and *Carex sempervirens* (D.SeCarS, $R = -0.78$). The negative scores correspond to areas with the opposite characteristics. The chamois habitat is the narrowest on this dimension of the ecological space, regarding the width of the distribution of available points. The factorial map of the ecological space indicates that the distribution of the available environment is shifted to the negative values of the first axis (whereas the used points are still centered on zero; see Fig. 3F). Within the studied context, it seems that the chamois select the areas close to the screes (50% of the detections within 111 m of this environment type) and, above all, close to meadows dominated by *Sesleria* and *Carex sempervirens* (75% of the detection within 70 m of this vegetation type).

The environmental suitability maps (ESMs) built using the first axis confirmed these results (Fig. 3B). The comparison of the ESMs with the maps of environmental variables showed that the most suitable areas are found close to meadows dominated by *Sesleria* and *Carex sempervirens*, and close to screes (Fig. 2C). The effect of the elevation here seems indirect: the most suitable areas are found at high elevation, which correspond to low distances to meadows dominated by *Sesleria* and *Carex sempervirens* (this environment type is on average located at an elevation of 1588 ± 183 m [mean ± SD]) and to screes (which were, on average, located at an elevation of 1748 ± 163 m). The indirect effect is consistent with the sharp aspect of the map that indicates a clear frontier between suitable and unsuitable environments, whereas the elevation map is more continuous. Note that the main spatial structures of the map of the full-rank Mahalanobis distances (Fig. 3C) are clearer on the ESMs built from the analysis (Fig. 3B): the increased precision (reduced generality) of the full-rank Mahalanobis distances is manifest in the identification of less area as potentially suitable (more noise is included in this measure).

Female chamois give birth to young in May and need a lot of resources to feed them (Hamr 1985, Ferrari et al. 1988). The prolific regrowth of the vegetation results in many energetic shoots in the meadows dominated by *Sesleria* and *Carex sempervirens*, which may therefore explain the abundance of the chamois in such environments at this time of the year. The distance to screes is also well-correlated with the first axis of the MADIFA, but this probably results from a confounding effect, as the screes are close to such meadows. This proximity of the screes probably increases the chamois preference for these meadows, as the screes may provide both an escape in case of predators (Bleich et al. 1997) and saline resource.

We then measured the goodness of fit with the validation data set. Following, Knick and Dyer (1997), we computed the cumulative frequency of the reduced-rank Mahalanobis distances (Fig. 3B) for (i) the pixels of the study area, (ii) the pixels containing chamois detections of the calibration set, and (iii) the pixels containing detections of the validation set (Fig. 4). We used the curves of both the study area and the validation set to derive a measure of the predictive capabilities of the analysis. The area located above the curve of the study area and below the curve of the validation set on this graph measures the quality of the prediction. Indeed, this area would be maximum in the case of a perfect prediction, because the value of the cumulative frequency of distances for the validation set would be equal to one whatever the value of distance (indicating that these distances are equal to zero for all the detections of the validation set). Therefore, dividing the quality of prediction of the validation set by the area located above the curve of the study area and below the line $Y = 1$ (theoretical perfect prediction) on this graph gives a standardized measure Q of quality of prediction. We also computed this ratio for the calibration data set, to give a measure G for the goodness of fit.

The goodness of fit of the MADIFA is rather high ($G = 74\%$; Fig. 4A). The curve of cumulative frequency distribution for the validation set is similar to the curve of the calibration set, indicating good predictive

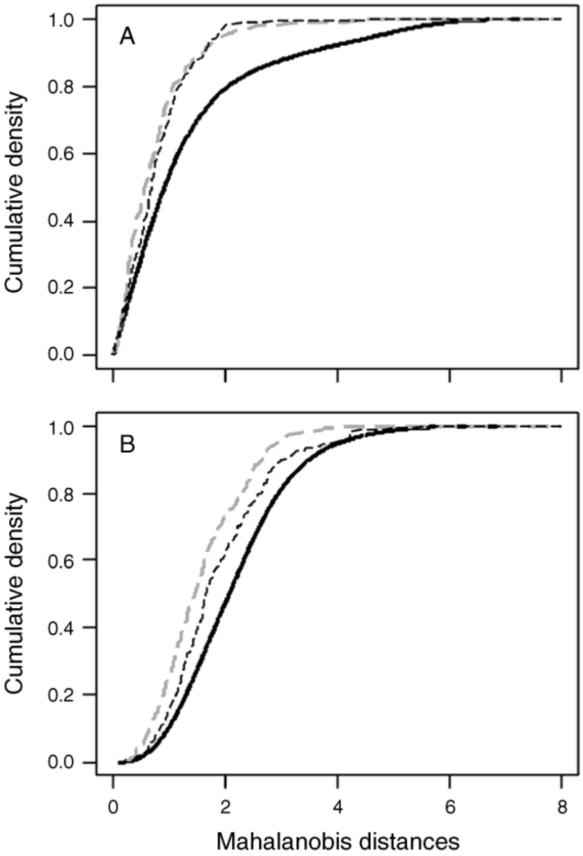

FIG. 4. Cumulative frequency distribution of the reduced-rank Mahalanobis distances computed for the pixels of the study area in the Bauges mountains (French Alps, solid black line), for the calibration data set (pixels where chamois groups were detected from 1994 to 2000, dashed gray line), and for the validation data set (pixels where chamois groups were detected from 2001 to 2004, dashed black line). (A) Reduced-rank Mahalanobis distances computed using the first axis of the MADIFA; and (B) reduced-rank Mahalanobis distances computed using the last seven axes of the PCA of the niche.

Plate 1. A chamois (*Rupricapra rupricapra*) photographed in the Bauges mountains (French Alps). Photo credit: Marc Cornillon.

capabilities ($Q = 73\%$). Indeed, 94% of the detections of these sets are in the top 75% of the reduced-rank Mahalanobis distances of the pixels of the study area.

Finally, we compared the results of the MADIFA with those of the PCA of the used points advocated by Rotenberry et al. (2002, 2006). We computed an ESM using the last seven axes of the PCA of the used points (Fig. 3D). The goodness of fit was lower than for the MADIFA ($G = 66\%$), and the predictive capabilities of this ESM were even lower ($Q = 59\%$, Fig. 4B). In fact, the main factor limiting the chamois distribution is closely related to the elevation, which is the main pattern on the study area. Therefore, this basic habitat requirement is unlikely to define the last axes of the PCA of the used points. Using the last axes of the PCA to build an ESM is likely to keep only the "noisy part" of the Mahalanobis distances. This again stresses the need to take into account the availability when one wants to identify habitat requirements.

Discussion

We developed Mahalanobis distance factor analysis (MADIFA) to explore, analyze, and visualize the niche in the ecological space. Furthermore, these results can be used to derive environmental suitability maps (ESMs) to visualize the patterns of the niche in the geographical space. This method led us to identify the main characteristics of the environment selected by the chamois, and provided an ESM of the area. We pointed out that the elevation is correlated to all the environmental variables included in the analysis (e.g., screes, meadows dominated by *Sesleria* and *Carex sempervirens* are generally found at high elevation) and is also the main structure of the chamois habitat: the variance of the species habitat is maximal for the elevation. However, although the chamois habitat is wider on this dimension, it is narrow relative to the range of available environment, indicating that this dimension contributes to the definition of a basic habitat requirement for this species (although indirectly, through its effect on the vegetation). This example clearly illustrates the need to take into account the availability in the partitioning of the Mahalanobis distances.

Accounting for the environmental availability at the time of sampling is also important for the "classical" Mahalanobis distances method. In most papers using this method, the environmental suitability is estimated on the area where the sample of used site has been collected (e.g., Clark et al. 1993). However, the environmental conditions may vary beyond the limits of this area. If the limits of the area on which the Mahalanobis distances are mapped are not carefully checked, the environmental conditions on the mapped area may not be representative of what was actually available to the species at the time of sampling. In such a case, the Mahalanobis distances may indicate an unsuitable environment in areas where the environmental conditions vary in a biologically positive direction (Knick and Rotenberry 1998). Consequently, even if the Mahalanobis distances method is a powerful method for ESM modeling, it does not circumvent the problem of the definition of availability.

Hypotheses underlying the MADIFA

The main assumption underlying the MADIFA is that the maximized statistic θ_j is relevant to capture the patterns of the niche in its environment. Because this statistic is a ratio between two sums of squared deviations from the mean of the niche, this assumption will be met if the mean of the niche is close to its mode (i.e., unimodal and symmetric niche). This hypothesis is also required by all factorial methods relying on the concept of ecological niche (ter Braak 1985, 1986, Knick and Rotenberry 1998, Hirzel et al. 2002). It ensures that the sum of squared deviations from the mean of the niche is a measure of the distance from the conditions most frequently used by the species.

This sum of squared deviations is very sensitive to outliers (Cleveland 1993), and so is the optimality criterion θ_j. Although this criterion allows MADIFA to be placed in a consistent theoretical framework (including the ecological-niche factor analysis [ENFA] and the Mahalanobis distances), further research needs to be done on factor analyses relying on more robust criteria, for example based on the median of absolute deviations from the median of the niche (Cleveland 1993).

Finally, one of the main issues regarding the statistical analysis of this type of data (therefore including MADIFA) is that most of the time the sample is not obtained using proper sampling designs that lead to unbiased estimation (e.g., random sampling or systematic sampling). The data concerning the chamois in the mountains of Les Bauges were obtained after a complete, therefore unbiased, census of the population in open areas, so that we did not meet this kind of problem. However, such sources of bias should be carefully checked in studies carried out at very large scale, especially in biogeography, where proper sampling is not possible (e.g., Spichiger et al. 2004).

Conclusions

The MADIFA is to be used jointly with other exploratory methods to visualize the structures of the niche. Classical PCAs can be used to identify correlates between environmental variables both in the species niche and on the study area. The MADIFA returns an image of the ecological space, and also allows visualization of the niche patterns in the geographical space, through the computation of an environmental suitability map (ESM). The ENFA may, in addition, be used to distinguish the parts of the Mahalanobis distances caused by the specialization and the marginality of the species. By matching all these results and the results of simpler descriptive statistics (e.g., histograms), the researcher can build a conceptual model of the biological system under study. The understanding of this system may be of major use for the estimation of more complex predictive models.

Acknowledgments

We warmly thank the Office National de la Chasse et de la Faune Sauvage for their financial support, and the Parc Naturel Régional des Bauges for providing the maps of environmental variables. We also thank two anonymous referees for their very constructive comments, which greatly improved the quality of this paper.

Literature Cited

Araujo, M. B., and P. H. Williams. 2000. Selecting areas for species persistence using occurrence data. Biological Conservation 96:331–345.

Basille, M., C. Calenge, E. Marboutin, R. Andersen, and J. M. Gaillard. 2008. Assessing habitat selection using multivariate statistics: some refinements of the ecological niche factor analysis. Ecological Modelling, *in press*.

Bleich, V. C., R. T. Bowyer, and J. D. Wehausen. 1997. Sexual segregation in mountain sheep: resources or predation? Wildlife Monographs 134:1–50.

Browning, D. M., S. J. Beaupré, and L. Duncan. 2005. Using partitioned Mahalanobis $D^2(k)$ to formulate a GIS-based model of timber rattlesnake hibernacula. Journal of Wildlife Management 69:33–44.

Bryan, T. L., and A. Metaxas. 2007. Predicting suitable habitat for deep-water gorgonian corals on the Atlantic and Pacific continental margins of North America. Marine Ecology Progress Series 330:113–126.

Calenge, C. 2006. The package adehabitat for the R software: a tool for the analysis of space and habitat use by animals. Ecological Modelling 197:516–519.

Cayuela, L. 2004. Habitat evaluation for the Iberian wolf *Canis lupus* in Picos de Europa National Park, Spain. Applied Geography 24:199–215.

Clark, J. D., J. E. Dunn, and K. G. Smith. 1993. A multivariate model of female black bear habitat use for a geographic information system. Journal of Wildlife Management 57: 519–526.

Cleveland, W. S. 1993. Visualizing data. Hobart Press, Summit, New Jersey, USA.

Corsi, F., E. Duprè, and L. Boitani. 1999. A large scale model of wolf distribution in Italy for conservation planning. Conservation Biology 13:150–159.

Dunn, J. E., and L. Duncan. 2000. Partitioning Mahalanobis D^2 to sharpen GIS classification. Pages 195–204 *in* C. A. Brebbia and P. Pascolo, editors. Management information systems 2000: GIS and remote sensing. WIT Press, Southampton, UK.

Elith, J., et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29: 129–151.

Farber, O., and R. Kadmon. 2003. Assessment of alternative approaches for bioclimatic modeling with special emphasis on the Mahalanobis distance. Ecological Modelling 160:115–130.

Ferrari, C., G. Rossi, and C. Cavani. 1988. Summer food habits and quality of female, kid and subadult Apennine chamois, *Rupricapra pyrenaica ornata* Neumann, 1899 (*Artiodactyla*, *Bovidae*). Z Säugetier 53:170–177.

Garcia-Gonzalez, R., and P. Cuartas. 1996. Trophic utilization of a montane/subalpine forest by chamois (*Rupicapra pyrenaica*) in the Central Pyrenees. Forest Ecology and Management 88:15–23.

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

Hamr, J. 1985. Seasonal home range size and utilisation by female chamois (*Rupicapra rupicapra*) in Northern Tyrol. Pages 106–116 *in* S. Lovari, editor. The biology and management of mountain ungulates. Croom Helm, London, UK.

Harville, D. A. 1997. Matrix algebra from a statistician's perspective. Springer, New York, New York, USA.

Hirzel, A. H., and R. Arlettaz. 2003. Modeling habitat suitability for complex species distributions by environmental-distance geometric mean. Environmental Management 32:614–623.

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.

Houssin, H., A. Loison, J. M. Gaillard, and J. M. Jullien. 1994. Validité d'une méthode d'estimation des effectifs de chamois dans un massif des préalpes du nord. Gibier Faune Sauvage 11:287–298.

Knick, S. T., and D. L. Dyer. 1997. Distribution of black-tailed jackrabbit habitat determined by GIS in southwestern Idaho. Journal of Wildlife Management 61:75–85.

Knick, S. T., and J. T. Rotenberry. 1998. Limitations to mapping habitat use areas in changing landscapes using the Mahalanobis distance statistic. Journal of Agricultural, Biological, and Environmental Statistics 3:311–322.

Legendre, P., and L. Legendre. 1998. Numerical ecology. Second English edition. Elsevier Science BV, Amsterdam, The Netherlands.

Mahalanobis, P. C. 1948. Historic note on the D^2 statistic. Sankhya 9:237–240.

Mäki-Petäys, A., T. Muotka, A. Huusko, P. Tikkanen, and P. Kreivi. 1997. Seasonal changes in habitat use and preference by juvenile brown trout, *Salmo trutta*, in a northern boreal river. Canadian Journal of Fisheries and Aquatic Sciences 54: 520–530.

Manly, B. F. J., L. L. McDonald, D. L. Thomas, T. L. MacDonald, and W. P. Erickson. 2002. Resource selection by animals. Statistical design and analysis for field studies. Kluwer Academic Publisher, London, UK.

R Development Core Team. 2005. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rameau, J. C., D. Mansion, and G. Dume. 2001. Flore forestière française, guide écologique, tome 2: montagnes. IDF, Paris, France.

Rotenberry, J. T., S. T. Knick, and J. E. Dunn. 2002. A minimalist approach to mapping species habitat: Pearson's planes of closest fit. Pages 281–289 *in* J. M. Scott, editor. Predicting species occurrence: issues of scale and accuracy. Island Press, Snowbird, Utah, USA.

Rotenberry, J. T., K. L. Preston, and S. T. Knick. 2006. GIS-based niche modeling for mapping species habitat. Ecology 87:1458–1464.

Spichiger, R., C. Calenge, and B. Bise. 2004. The geographical zonation in the Neotropics of tree species characteristic of the Paraguay-Paraná Basin. Journal of Biogeography 31:1489–1501.

ter Braak, C. J. F. 1985. Correspondence analysis of incidence and abundance data: properties in terms of a unimodal response model. Biometrics 41:859–873.

ter Braak, C. J. F. 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology 67:1167–1179.

Thompson, L. M., F. T. van Manen, S. E. Schlarbaum, and M. DePoy. 2006. A spatial modeling approach to identify potential butternut restoration sites in Mammoth Cave National Park. Restoration Ecology 14:298–296.

von Elsner-Schack, I. V. 1985. What is good chamois habitat? Pages 71–76 *in* S. Lovari, editor. The biology and management of mountain ungulates. Croom Helm, London, UK.

APPENDIX A

Mahalanobis distances (*Ecological Archives* E089-030-A1).

APPENDIX B

Demonstration: the sum of the squared scores of the pixels on the factorial axes of the MADIFA is equal to the Mahalanobis distances (*Ecological Archives* E089-030-A2).

APPENDIX C

Results of the principal component analyses performed to identify the correlations on the study area, and in the chamois niche (*Ecological Archives* E089-030-A3).