Journal of Theoretical Biology 252 (2008) 674–685

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

A general framework for the statistical exploration of the ecological niche

Clément Calenge [a,b,*], Mathieu Basille [a,c]

[a] Université de Lyon, F-69000, Lyon, France; Université Lyon 1, CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622, Villeurbanne, France
[b] Office national de la chasse et de la faune sauvage, 5, rue de Saint Thibault, Saint Benoist, 78610, Auffargis, France
[c] Centre for Conservation Biology, Department of Biology, Norwegian University of Science and Technology, NO-7491, Trondheim, Norway

ARTICLE INFO

Article history:
Received 18 September 2007
Received in revised form
9 February 2008
Accepted 27 February 2008
Available online 3 March 2008

Keywords:
Multivariate analysis
Factor analysis
Inertia ratio
Habitat selection
General niche-environment system factor analysis

ABSTRACT

We propose a new statistical framework for the exploratory analysis of the ecological niche, the "General niche-environment system factor analysis" (GNESFA). The data required for this analysis are (i) a table giving the values of the environmental variables in each environment unit (EU, e.g., the patches of habitat on a vector map), (ii) a set of weights measuring the availability of the EUs to the species (e.g., the proportion of the study area covered by a given patch), and (iii) a set of utilization weights describing the use of the EUs by the focal species (e.g., the proportion of detections of the species in each patch). Each row of the table corresponds to a point in the multidimensional space defined by the environmental variables, and each point is associated with two weights. The GNESFA searches the directions in this space where the two weight distributions differ the most, choosing one distribution as the reference, and the other one as the focus. The choice of the utilization as the reference corresponds to the MADIFA (Mahalanobis distances factor analysis), which identifies the directions on which the available EUs are in average the furthest from the optimum of the niche, allowing habitat suitability modelling. The choice of the availability as the reference corresponds to the FANTER (Factor analysis of the niche, taking the environment as the reference), which identifies the directions on which the niche is the furthest from the average environment (marginality) and those on which the niche is the narrowest compared with the environment (specialization). The commonly used ENFA (Ecological niche factor analysis) is at the middle point between the MADIFA and the FANTER, considering both distributions as the reference and the focus simultaneously. When used concurrently, these three analyses allow an extensive exploration of the system.

1. Introduction

Two kinds of studies can be carried out to study the relationships between a species and its environment (Morrison et al., 1992; Calenge et al., 2005). On one hand, hindcasting studies aim to emphasize among a large set of environmental variables those that are of interest for the species. On the other hand, forecasting modelling is used to predict suitable environments for the species in new unsampled areas, and/or under different environmental conditions (Guisan et al., 2006; Knick and Rotenberry, 1998). With the predicted global warming of the climate, it is of increasing importance to predict the behavior of keystone species in response to various scenarios of future climate (e.g., Araujo et al., 2005). For this reason, the social and scientific demand for predictive models is presently very strong (Elith et al., 2006). However, statistical methods allowing forecasting modelling, such as general linear model or related methods, can deal only with a limited number of environmental variables (Guisan and Zimmermann, 2000). For this reason, it is generally supposed in such studies that "the modeller knows the limiting factors that influence the distribution and abundance of the study organism" (Boyce and McDonald, 1999). All these methods rely on the hypothesis that a large amount of biological knowledge concerning the species is available prior to the study (Burnham and Anderson, 1998). In other words, hindcasting studies should necessarily precede a forecasting approach (Soberon and Peterson, 2005).

Visualization is an essential step of hindcasting studies. As noted by Cleveland (1993), "visualization is critical to data analysis. It provides a front line of attack, revealing intricate structure in data that cannot be absorbed in any other way. We discover unimagined effects, and we challenge imagined ones". In particular, only the use of a visualization approach would reveal hidden structures and other "surprises" in the data (Cleveland, 1994).

---

* Corresponding author at: Université Lyon 1, CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622, Villeurbanne, France.
Tel.: +33 478431357; fax: +33 478431388.
E-mail address: calenge@biomserv.univ-lyon1.fr (C. Calenge).

0022-5193/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2008.02.036

The graphical exploration of the relationships between a species and its environment may rely on the formal concept of ecological niche (Hutchinson, 1957). Each environmental variable can define a dimension of a multidimensional space, namely the ecological space. In that space, the distribution of the species occurrences represents the niche, which can be compared to the environment defined as available to the species (e.g., pixels of a raster map). This concept allows both a graphical and a quantitative exploratory analysis, in order to identify the directions in the ecological space where the distribution of the species is most different from the distribution of points describing the environment available to the species. However, the present "lack of effective tools for exploring, analysing, and visualizing ecological niches in many-dimensional environmental space" (Soberon and Peterson, 2005) may render this task difficult.

Factor analyses have numerous desirable qualities for data exploration in highly multidimensional spaces, especially for visualizing structures in the data (e.g., Hill, 1974; Blondel et al., 1988; Thioulouse and Chessel, 1992; Dray et al., 2003). For this reason, factor analyses may have a key role in hindcasting studies of species-environment relationships. They could be used to sort factorial axes which support most of the difference between the niche and its environment. The patterns of the niche-environment system identified by the analysis can then be related to the choices (habitat selection) or the requirements (niche characteristics) of the species concerning its habitat.

The Ecological-niche factor analysis (ENFA, Hirzel et al., 2002) and the Mahalanobis distances factor analysis (MADIFA, Calenge et al., 2008) are two such methods. On one hand, the ENFA distinguishes two kinds of information measured in the niche-environment system, marginality and specialization. The marginality is a measure of the eccentricity of the niche relative to the distribution of available points in the ecological space, whereas the specialization is a measure of the niche restriction relative to the distribution of available points. The ENFA comes up with the directions in the ecological space where first the marginality, and then the specialization are maximized. On the other hand, the MADIFA relies only on one measure of habitat suitability, Mahalanobis distance, computed in the ecological space from the centroid of the distribution of the species occurrences to all available points. The Mahalanobis distance gives an index of the environmental suitability at this place, as it indicates the departure from the species' optimum (a low distance value is supposed to indicate a strong suitability, see Clark et al., 1993; Knick and Dyer, 1997). The MADIFA returns the directions in the ecological space where the Mahalanobis distances of the available sites are, on average, the largest. Both the ENFA and the MADIFA have been proposed as appropriate tools for drawing factorial maps of the niche in the ecological space, or for building reduced-rank environmental suitability maps (Hirzel et al., 2002; Calenge et al., 2008).

From a formal point of view, the ENFA and MADIFA are actually closely related (Calenge et al., 2008). In this paper, we extend the mathematical relationships between the ENFA and the MADIFA to develop a more general framework encompassing these two methods, the general Niche-environment system factor analysis (GNESFA). This framework also includes another factor analysis of the niche-environment system, the Factor analysis of the niche, taking the environment as the reference (FANTER), which offers a third point of view regarding this system. These three exploratory methods return complementary results, as illustrated by two examples. When used concurrently, they provide an extensive summary of the patterns in the data.

2. The algorithm

2.1. Notations and definitions

We suppose that the study area is made of a set of N discrete environment units (EU), on which P environmental variables are measured (Fig. 1). These EUs may be, for example, the pixels of a raster map, or the patches of environment on a vector map. These values are stored in the $(N \times P)$ matrix $\mathbf{X}$. Because the environmental variables may not be measured using the same units (e.g., the elevation is measured in meters, and the slope in percent), we suppose that the columns of $\mathbf{X}$ have been standardized prior to the analysis (with zero mean and unit variance).

A weight describes the availability of each EU to the focal species. This "availability weight" could be, for example, the proportion of the study area covered by a patch of environment (in the case of raster maps, all these availability weights may be equal). This set of weights—defining the "availability distribution" —is stored on the diagonal of the $(N \times N)$ matrix $\mathbf{D}_A$ (the values of the non-diagonal elements of $\mathbf{D}_A$ are set to 0).

Additionally, an "utilization weight" describes the intensity of use of each EU by the focal species. This weight could be, for example, the proportion of detections of the species located in the patch during a census operation. The set of utilization weights—defining the "utilization distribution"—is stored on the diagonal of the $(N \times N)$ matrix $\mathbf{D}_U$ (the values of the non-diagonal elements of $\mathbf{D}_U$ are set to 0). Note that both the utilization and availability weights sum to 1 by construction. The aim of the GNESFA is to identify the directions in the ecological space where the two weight distributions differ most, which we call "niche patterns".

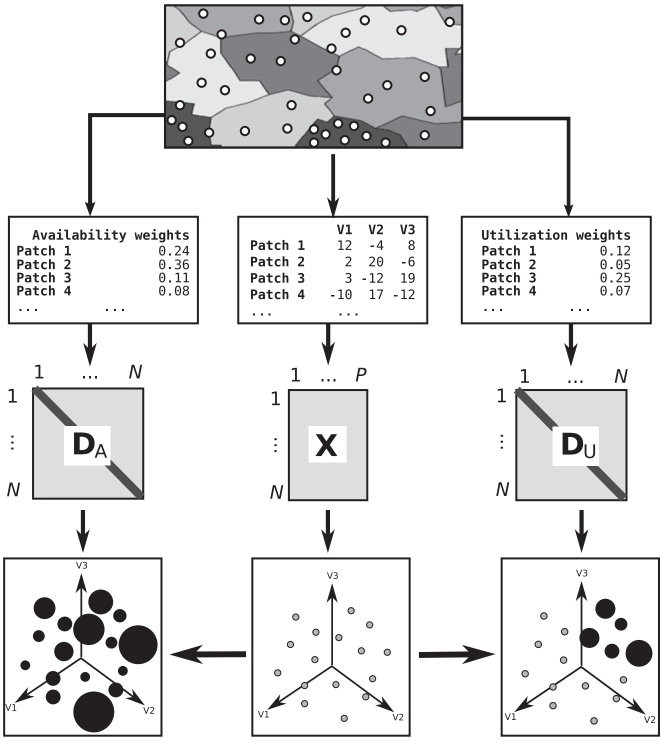

Fig. 1. The data design required by the GNESFA: the study area is partitioned into a set of N discrete environment units (here, patches of environments). For each patch, P environmental variables are measured (here, V1, V2, and V3 and stored in the matrix $\mathbf{X}$). Each variable defines a dimension of the ecological space. For each environment unit, an "availability weight" (stored in the diagonal matrix $\mathbf{D}_A$) defines its availability to the species, and an "utilization weight" (stored in the diagonal matrix $\mathbf{D}_U$) measures its intensity of use by the species.

2.2. Choice of a reference

The GNESFA implies a choice of one of these two weight distributions as a reference distribution, and the other as a focus distribution (Fig. 2). The cloud of points defined by the rows of **X** will be distorted so that this cloud, considered from the point of view of the reference distribution, will take a standard spherical shape (i.e., with a variance of the available EUs equal to one for all directions of the ecological space). Then, the shape of the cloud of points considered from the point of view of the focus distribution will be investigated in this standard space, and any deviation from this spherical shape will indicate a pattern. The choice of a reference distribution depends on the needs of the analyst.

On one hand, when the main interest of the analysis is related to the identification of the variables affecting the shape (unimodal vs multimodal niches), the central tendency (marginal species or not), and the spread of the niche relative to the environment (specialized species or not), the availability distribution should be chosen as the reference, and the utilization distribution as the focus. Such a choice implies that the patterns of the available EUs are known (i.e., the correlation structure of the environmental variables on the study area), and would correspond to the point of view of the expert of the study area. Therefore, the choice of the availability as a reference allows a detailed exploratory analysis of the patterns displayed by the niche in the ecological space. This is the case of the Factor analysis of the niche, taking the environment as the reference (FANTER, Fig. 2).

On the other hand, in some cases, we are more interested in the patterns of the environment relative to the niche. For example, the suitability of the available EUs can be measured by the distance between these EUs and the utilization distribution as a whole (Clark et al., 1993). It follows that, if one wants to know the most suitable conditions of an area considered as available to the species, the utilization distribution should be chosen as the reference. The distribution of used EUs will then take a standard shape, and the GNESFA will indicate the direction of the ecological space in which the available EUs are the most different from this distribution. Such a choice implies that the patterns of the utilization distribution are known, and corresponds to the point of view of the expert of the studied species. In that context, we are mainly interested in whether the species "considers" the proposed environment (available EUs) as suitable (within the niche) or not (far from the niche). This is the approach used by the Mahalanobis distances factor analysis (MADIFA, Fig. 2).

Finally, another point of view is possible, for which each of the two distributions are both the reference and the focus distribution. This symmetrical point of view has the advantage of not relying on the choice of one distribution as the reference. However, we will later prove that this choice also implies the loss of one dimension of the ecological space, a dimension which may carry important biological information. This special case is the basis of the Ecological-niche factor analysis (ENFA, Fig. 2).

These three approaches define the framework of the GNESFA that we describe now more formally. The reference distribution is described by the weight matrix **R**, and the focus distribution is described by the weight matrix **F**. For example, if the availability distribution is chosen as reference, then $\mathbf{R} = \mathbf{D}_A$ and $\mathbf{F} = \mathbf{D}_U$. In this paper, we will refer to the "reference mean", "reference variance", "focus mean" or "focus variance" of a variable, depending on the computed statistic (mean or variance) and the weight distribution chosen to compute this statistic (**R** and **F**, respectively). We call the "centroid of the reference distribution" the point in the ecological space defined by the vector $\mathbf{X}^t\mathbf{R}\mathbf{1}_N$ (i.e., the P-vector of reference means computed for all the environmental variables). Similarly, we will refer to the "centroid of the focus distribution" if the chosen weight matrix is the matrix **F**.

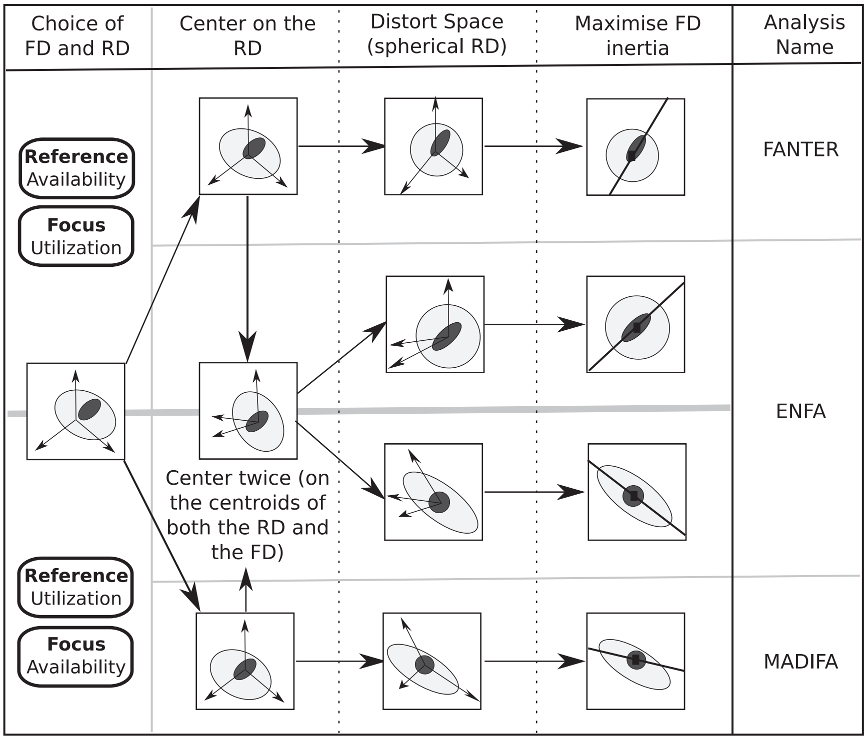

Fig. 2. The three possible analyses encompassed by the GNESFA. The light grey ellipse represents the distribution of availability weights in the ecological space and the dark grey ellipse represents the distribution of utilization weights in the ecological space. The FANTER uses the availability distribution as reference and the utilization distribution as focus. The MADIFA uses the utilization distribution as reference and the availability distribution as focus. The ENFA can use both approaches (RD = Reference distribution; FD = Focus distribution).

2.3. *The centering*

The first step of the GNESFA is the centering of the table **X**. Actually, “mathematically and geometrically, centering involves the specification of the origin (...). It is the ‘point of zero information’; anything that is at it, is trivial and uninteresting; anything that deviates from it is information” (Noy-Meir, 1973). Therefore, it seems logical to choose the centroid of the reference distribution as the origin of the ecological space to perform the GNESFA. The centered table **Z** is thus computed by (Seber, 1984)

$$\mathbf{Z} = (\mathbf{I}_N - \mathbf{1}_N \mathbf{1}_N^{\mathrm{t}} \mathbf{R})\mathbf{X}$$

where $\mathbf{I}_N$ is the $N \times N$ identity matrix, and $\mathbf{1}_N$ is a N-vector of 1s.

2.4. *First principal component analysis*

The second step of the GNESFA consists of a principal component analysis (PCA) of the table **Z**, using the matrix **R** as row weights. Actually, the PCA of **Z** consists of the search for the eigenstructure of the variance–covariance matrix **V**:

$$\mathbf{V} = \mathbf{Z}^{\mathrm{t}}\mathbf{R}\mathbf{Z}$$

Let **U** be the $(N \times P)$ matrix containing the P eigenvectors $\mathbf{u}_j$ of **V** concatenated by column, and let $\boldsymbol{\Lambda}$ be the diagonal matrix containing the corresponding eigenvalues λ_j, on the diagonal. In other words:

$$\mathbf{V}\mathbf{U} = \mathbf{U}\boldsymbol{\Lambda}$$

The $(P \times P)$ matrix **U** contains the scores of environmental variables (rows) on each principal axis of the analysis (columns). Moreover, the $(N \times P)$ matrix $\mathbf{L}^{\bullet} = \mathbf{Z}\mathbf{U}$ contains the coordinates of the EUs (rows) on the principal components of the analysis (columns) (Legendre and Legendre, 1998).

Because the table **Z** is centered for the weight matrix **R**, this analysis is just a classical PCA, i.e., a rotation of the cloud of points so that: (i) the reference variance of the EU coordinates on the first principal components is maximized (it is equal to the corresponding eigenvalues), and (ii) the correlation between the coordinates of the EUs on different principal components is equal to 0 (Legendre and Legendre, 1998). Let the $(N \times P)$ matrix **L** be equal to

$$\mathbf{L} = \mathbf{L}^{\bullet}\boldsymbol{\Lambda}^{-1/2} = \mathbf{Z}\mathbf{U}\boldsymbol{\Lambda}^{-1/2}$$

This matrix contains the normed coordinates of the EUs (rows) on the principal components (columns) of the analysis. This matrix has the following property:

$$\mathbf{L}^{\mathrm{t}}\mathbf{R}\mathbf{L} = \mathbf{I}_P \tag{1}$$

where $\mathbf{I}_P$ is the $P \times P$ identify matrix. This equation shows that the product of the EU coordinates and $\boldsymbol{\Lambda}^{-1/2}$ results in a distortion of the cloud of EUs in the ecological space, so that the reference variance of these coordinates after distortion is equal to 1 for all the principal components. As these components are still uncorrelated, it follows that the cloud of points described by **X** has been “sphericized”, from the point of view of the reference distribution.

2.5. *Second principal component analysis: the core of the GNESFA*

The last step of the GNESFA is the analysis of the focus distribution in this distorted ecological space. This analysis is done using a non-centered PCA of the table **L**, with the matrix **F** as row weights.

Since the cloud of points has been “sphericized” with respect to the reference distribution, it should also be spherical from the point of view of the focus distribution, if it is identical to the reference distribution. In other words, all the eigenvalues of this PCA should be equal, which would indicate that: (i) the centroid of the focus distribution is the same as the centroid of the availability distribution (because the analysis is not centered), and (ii) the variance of the focus distribution is the same in all the directions of the ecological space. More formal justifications will be given in the next section.

This PCA is performed by computing the eigenstructure of the matrix **H**:

$$\mathbf{H} = \mathbf{L}^{\mathrm{t}}\mathbf{F}\mathbf{L}$$

This PCA is non-centered because **L** is not centered for the weight matrix **F**. Let **W** be the matrix $P \times P$ containing the eigenvectors $\mathbf{w}_j$ of **H** concatenated by column, and the diagonal matrix $\boldsymbol{\Gamma}$ containing the eigenvalues γ_j of **H** on the diagonal:

$$\mathbf{H}\mathbf{W} = \mathbf{W}\boldsymbol{\Gamma}$$

The coordinates of the EUs on the principal components of the GNESFA are contained in the $N \times P$ matrix **P**:

$$\mathbf{P} = \mathbf{L}\mathbf{W} = \mathbf{Z}\mathbf{U}\boldsymbol{\Lambda}^{-1/2}\mathbf{W}$$

This equation summarizes the GNESFA: it consists of a first rotation (matrix **U**), a distortion (matrix $\boldsymbol{\Lambda}^{-1/2}$) and a second rotation (matrix **W**) of the cloud of points in the ecological space (matrix **Z**).

Note that the cloud of points is still spherical with respect to the reference distribution on the components of the GNESFA:

$$\mathbf{P}^{\mathrm{t}}\mathbf{R}\mathbf{P} = \mathbf{W}^{\mathrm{t}}\mathbf{L}^{\mathrm{t}}\mathbf{R}\mathbf{L}\mathbf{W} = \mathbf{I}_P \tag{2}$$

This arises from Eq. (1) and from the observation that **W** is an orthogonal matrix (i.e., $\mathbf{W}^{\mathrm{t}}\mathbf{W} = \mathbf{W}\mathbf{W}^{\mathrm{t}} = \mathbf{I}_P$).

Factorial maps of the niche in the ecological space can be obtained by plotting the coordinates of the EUs on a restricted number of principal components, as in classical PCA. The biological meaning of the principal components can be derived from the correlations between the environmental variables and the principal components of the analysis. Note that some school of thought prefer to interpret the meaning of the principal components from the scores of the environmental variables on the principal axes of the GNESFA, contained in the matrix **A** (Rotenberry et al., 2006)

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}^{-1/2}\mathbf{W}$$

We advocate the use of the correlations to interpret the meaning of the principal components. Indeed, the coefficient associated with a given environmental variable may be misleading when this variable is correlated with other environmental variables in **X** (Basille et al., 2008).

To choose the number of components to keep for the interpretation, we can look for a break in the decrease of the eigenvalues (broken-stick method, Barton and David, 1956; Frontier, 1976). The biological meaning of these eigenvalues depends on the weight matrix chosen as reference distribution, as developed in the next section.

2.6. *The inertia ratio*

We now investigate the mathematical meaning of the statistic maximized by the GNESFA on the first components of the analysis. Because the second step of the GNESFA is a PCA, the jth eigenvalue γ_j of this analysis is equal to

$$\gamma_j = \mathbf{p}_j^{\mathrm{t}}\mathbf{F}\mathbf{p}_j \tag{3}$$

where $\mathbf{p}_j$ is the jth column of **P** (i.e., the jth component of the analysis). Note that, as in classical PCA, this statistic is at its maximum on the first axis of the analysis. For this reason, we focus on this first axis to make explicit the properties of this statistic. The first axis of the GNESFA $\mathbf{a}_1$ (i.e., the first column of $\mathbf{A}$) fulfills the following conditions:

$$\mathbf{Z}\mathbf{a}_1 = \mathbf{p}_1 \tag{4}$$

$$\mathbf{p}_1^{t}\mathbf{R}\mathbf{p}_1 = 1 \tag{5}$$

$$\gamma_1 = \mathbf{p}_1^{t}\mathbf{F}\mathbf{p}_1 \quad \text{Max} \tag{6}$$

The condition 4 means that the vector $\mathbf{p}_1$ contains a linear combination of the environmental variables, using the values stored in $\mathbf{a}_1$ as coefficients. This linear combination gives the coordinates of the EUs on the principal components of the analysis, such that: (i) the reference variance of these coordinates is equal to 1 (condition 5, arising from Eq. (2), and (ii) the focus mean of squared coordinates is maximized (condition 6, arising from Eq. (3)).

Actually, we can show that the conditions 4–6 define a problem mathematically equivalent to the search for a *P*-vector $\mathbf{g}_1$, fulfilling the following conditions (see Appendix A):

$$\mathbf{g}_1^{t}\mathbf{g}_1 = 1 \tag{7}$$

$$\mathbf{Z}\mathbf{g}_1 = \mathbf{y} \tag{8}$$

$$\gamma_1 = \frac{\mathbf{y}^{t}\mathbf{F}\mathbf{y}}{\mathbf{y}^{t}\mathbf{R}\mathbf{y}} \quad \text{Max} \tag{9}$$

The vector $\mathbf{g}_1$ is normed to 1 (condition 7). Consequently, the vector $\mathbf{y}$ contains the coordinates of the orthogonal projections of the undistorted cloud of EUs in the ecological space on the vector $\mathbf{g}_1$ (condition 8). The coordinates of these projections are such that the ratio γ_1 is maximized (condition 9). This equivalence between the two problems is demonstrated in Appendix A, with

$$\mathbf{g}_1 = \frac{\mathbf{a}_1}{\sqrt{\mathbf{a}_1^{t}\mathbf{a}_1}}$$

Therefore, the first eigenvalue γ_1 of the GNESFA is equal to

$$\gamma_1 = \frac{\sum_{i=1}^{n} f_i (y_i - \bar{y}_r)^2}{\sum_{i=1}^{n} r_i (y_i - \bar{y}_r)^2}$$

where y_i is the ith element of the vector $\mathbf{y}$, $\bar{y}_r$ is the reference mean of $\mathbf{y}$, r_i is the ith reference weight, and f_i is the ith focus weight.

The denominator of γ_j is the reference variance of $\mathbf{y}$. On the contrary, the numerator is generally not a variance, except if the focus mean of $\mathbf{y}$ is equal to its reference mean: it is the focus mean of the squared distances between the EUs and the centroid of the reference distribution. Such a sum of squared distances is sometimes called "inertia" (Dray et al., 2003). For this reason, we named γ_j the "inertia ratio" of the jth component of the GNESFA. The biological meaning of this ratio depends on the distribution chosen as the reference (see below).

3. Three special cases

As explained before, the GNESFA encompasses three methods, depending on the choice made by the analyst for the focus and reference distributions (Fig. 2). When the utilization distribution is chosen as reference, the resulting analysis is the MADIFA (Calenge et al., 2008). When the availability distribution is chosen as reference, the resulting analysis is a new analysis, which we called the Factor analysis of the niche, taking the environment as the reference (FANTER). Finally, the ENFA (Hirzel et al., 2002) is also a special case of the GNESFA, provided that a special transformation of the table $\mathbf{X}$ has been carried out prior to the analysis. Because the point of view of the ENFA is central to the understanding of the differences between these three analyses, we first describe how the ENFA takes place within the framework of the GNESFA.

3.1. The ENFA: a prior transformation of the table

A classical approach for characterizing a statistical distribution is to provide a measure of its central tendency (e.g., mean, median) and a measure of its spread (e.g., variance, interquartile range). The ENFA relies on this approach, in a multidimensional space, to characterize the niche-environment system.

On one hand, the "central tendency" of the utilization distribution relative to the availability distribution is measured by the vector connecting the centroids of the two distributions, named the "marginality vector". Its computation can be performed in the following way: first the table $\mathbf{X}$ is centered for the weight matrix $\mathbf{D}_A$:

$$\mathbf{C} = (\mathbf{I}_N - \mathbf{1}_N\mathbf{1}_N^{t}\mathbf{D}_A)\mathbf{X}$$

Then the marginality vector $\mathbf{m}$ can be computed as the *P*-vector containing the utilization means of the columns of $\mathbf{C}$:

$$\mathbf{m} = \mathbf{C}^{t}\mathbf{D}_U\mathbf{1}_N$$

This vector contains the differences between the utilization and availability means of all the environmental variables. Its squared length (equal to $m^2 = \mathbf{m}^{t}\mathbf{m}$), named "marginality", is a measure of the eccentricity of the utilization distribution relative to the availability distribution.

On the other hand, the spread of the utilization relative to the availability distribution may vary according to the considered direction of the ecological space. Hirzel et al. (2002) proposed the use of the specialization ratio *S*:

$$S = \frac{\text{availability variance}}{\text{utilization variance}}$$

For a given environmental variable, a large *S* indicates that the environmental variability experienced by the species is much smaller than the range of variability that is actually available, and therefore that the species is highly specialized on this variable. The aim of the ENFA is to identify the directions, in the ecological space, where the specialization ratio *S* is maximized.

The equivalence of the ENFA and the GNESFA can be proved by noting that the specialization ratio *S* is a special case of the inertia ratio γ_j. We noted previously that the numerator of γ_j is a variance only if the centroid of the reference distribution is the same as the centroid of the focus distribution. Therefore, the only way to compute the specialization ratio is to consider only those directions of the ecological space orthogonal to the marginality vector. For these directions, both the availability means and the utilization means are equal to zero. The projection of the EUs on the hyperplane orthogonal to $\mathbf{m}$ is carried out by the following operation (Harville, 1997):

$$\mathbf{C}_c = \mathbf{C}\left(\mathbf{I}_v - \frac{\mathbf{m}\mathbf{m}^{t}}{\mathbf{m}^{t}\mathbf{m}}\right) \tag{10}$$

The GNESFA of the table $\mathbf{C}_c$, using $\mathbf{D}_U$ as the reference distribution, and $\mathbf{D}_A$ as a focus distribution, is the solution to the following problem:

$$\mathbf{g}_1^{t}\mathbf{m} = 0 \tag{11}$$

$$\mathbf{g}_1^{t}\mathbf{g}_1 = 1 \tag{12}$$

$$\mathbf{Z}\mathbf{g}_1 = \mathbf{y} \tag{13}$$

$$S = \frac{\mathbf{y}^{t}\mathbf{D}_A\mathbf{y}}{\mathbf{y}^{t}\mathbf{D}_U\mathbf{y}} \quad \text{Max} \tag{14}$$

The conditions 12 and 14 are identical to the conditions 7 and 9 previously defined. The condition 11 derives from equation 10 (as the EUs are all located on the hyperplane orthogonal to the marginality vector). The condition 13 derives from the condition 8 (i.e., in this case $\mathbf{C}_c\mathbf{g}_1 = \mathbf{y}$) and the condition 11. This problem is exactly the problem of the ENFA defined by Hirzel et al. (2002). The ENFA is therefore a special case of the GNESFA.

We can see that the first component of the GNESFA of the table $\mathbf{C}_c$ using $\mathbf{D}_A$ as reference distribution maximizes $1/S$. Consequently, this component is the same as the last component of the GNESFA of the table $\mathbf{C}_c$, using $\mathbf{D}_U$ as reference distribution. Thus, the GNESFA of the table $\mathbf{C}_c$ can be seen as an ENFA, whatever the weight distribution chosen as reference. When $\mathbf{D}_U$ is chosen as the reference, the GNESFA is the classical ENFA. When $\mathbf{D}_A$ is chosen as the reference, the GNESFA is a "reversed" ENFA (with the first axes of the classical ENFA being the last axes of the "reversed" ENFA). Therefore, the roles of the utilization and availability distributions are symmetric in the ENFA (see Fig. 2). This corresponds to the symmetrical point of view described previously.

Whatever the importance of the specialization, the dimension of the ecological space defined by the marginality vector may carry important information about the niche-environment pattern, and the biological meaning of this dimension should also be interpreted (Hirzel et al., 2002). Thus, the EUs are also projected on this vector, to define the marginality component $\mathbf{b}$:

$$\mathbf{b} = \mathbf{C}\frac{\mathbf{m}}{\sqrt{\mathbf{m}^t\mathbf{m}}}$$

and the values in $\mathbf{b}$ can be plotted vs. values in $\mathbf{y}$ (row coordinates on the specialization axes) to build factorial maps of the niche-environment system. Basille et al. (2008) have proved that such maps give an optimal image of the niche from the point of view of the ENFA (distinction between marginality and specialization), and are undistorted because the marginality vector is orthogonal to the specialization axes of the ENFA.

However, the ENFA may give problematic results in three cases: (i) for some data sets, the marginality is not biologically significant. In such cases, imposing the constraint of orthogonality of the specialization axes to the marginality vector may lead to meaningless results. Indeed, even if the marginality is not strong, the specialization may be important for the marginality component, and the orthogonality constraint may obscure this characteristic of the data; (ii) although this method has been widely used to build habitat suitability maps (e.g., Zaniewski et al., 2002; Reutter et al., 2003), we believe that this method should not be used to reach such a goal. Because the marginality component does not have the same mathematical status as the successive specialization components, it is very peculiar to combine all of them into a single value of habitat suitability. Even though the *ad hoc* methods existing in the literature (Hirzel and Arlettaz, 2003) have returned biologically consistent results, we believe that the ENFA is not designed to build such maps, and that better methods exist for this objective (see below); (iii) finally, the ENFA relies on the hypothesis that both the utilization and the availability distributions are symmetric and unimodal (Hirzel et al., 2002). In the case of a multimodal niche, the ENFA is not recommended (see examples below).

3.2. *The MADIFA: utilization distribution as reference*

The specialization and the marginality are clear measures of the niche patterns in the ecological space. We can express the inertia ratio γ_j as a function of the marginality and the specialization, in order to give insight into the differences between the ENFA, the MADIFA and the FANTER.

When the reference is the utilization distribution (i.e., $\mathbf{R} = \mathbf{D}_U$ and $\mathbf{F} = \mathbf{D}_A$), the ratio $\gamma_{(m)j}$ (subscript "*m*" stands for "MADIFA"), maximized by the analysis can be reformulated:

$$\gamma_{(m)j} = \frac{m_j^2}{s_{(u)j}^2} + S_j$$

where $s_{(u)j}^2$ is the utilization variance of the jth component, m_j^2 measures the marginality on this component (the squared difference between the availability mean and the utilization mean of this component), and S_j is the specialization on this component. Thus, this analysis combines the marginality and the specialization into one single value: the larger the inertia ratio, the higher the marginality and/or the specialization.

Calenge et al. (2008) has already described this special case of the GNESFA, and called it MADIFA (Mahalanobis distances factor analysis). The authors noted an interesting property of this analysis: the sum of squared scores of an available EU over all the components of the analysis is equal to the Mahalanobis distance between this EU and the centroid of the utilization distribution. More formally,

$$D_i^2 = \mathbf{P}_{i\bullet}\mathbf{P}_{i\bullet}^t = \sum_{j=1}^{P} p_{ij}^2$$

where the P-vector $\mathbf{P}_{i\bullet}$ is the ith row of the matrix $\mathbf{P}$, and p_{ij} is the coordinate of the ith EU on the jth component of the GNESFA. Thus, the squared coordinate of a EU on a given component of the GNESFA can be considered as the contribution of this component to the Mahalanobis distance between this EU and the centroid of the reference distribution. This property is interesting because the Mahalanobis distances have been used in many studies as indices of environmental suitability for species (Clark et al., 1993; Farber and Kadmon, 2003; Knick and Dyer, 1997; Cayuela, 2004; Browning et al., 2005; Corsi et al., 1999), especially to build environment suitability maps.

Recalling Eq. (3), the inertia ratio on the jth component of the GNESFA is equal to

$$\gamma_j = \mathbf{p}_j^t\mathbf{D}_A\mathbf{p}_j = \sum_{i=1}^{N} a_i p_{ij}^2$$

where a_i is the availability weight associated with the ith EU. It follows that the jth eigenvalue of the analysis corresponds to the availability mean of the contributions of the jth component to the Mahalanobis distances between the EUs and the centroid of the utilization distribution.

Therefore, the MADIFA finds the directions of the ecological space where these distances are the largest, corresponding to the environmental conditions that are scarcely used by the species. Because they all have the same mathematical status, these components can be easily combined to build reduced-rank environment suitability maps with increased generality (see Calenge et al., 2008, for details). These directions are those where the marginality and/or the specialization are the largest. Thus, this analysis identifies all the patterns of the niche-environment system on the first principal components.

Note that the MADIFA relies on the hypothesis that the utilization distribution is both unimodal and symmetric (Calenge et al., 2008) and is therefore not recommended for the study of multimodal niches (as the ENFA). However, this analysis does not rely on any assumption concerning the shape of the availability distribution.

3.3. *The FANTER: availability distribution as reference*

When the reference is the availability distribution (i.e., $\mathbf{R} = \mathbf{D}_A$ and $\mathbf{F} = \mathbf{D}_U$), the ratio $\gamma_{(f)j}$ (subscript "*f*" stands for "FANTER") maximized by the GNESFA can be reformulated:

$$\gamma_{(f)j} = \frac{1}{S_j} + \frac{m_j^2}{s_{(a)j}^2}$$

where $s_{(a)j}^2$ is the availability variance of the jth component. Consequently, a large $\gamma_{(r)j}$ may indicate that the marginality is large, but also that the specialization is low. On the other hand, a low $\gamma_{(r)j}$ indicates a strong specialization and/or a low marginality. Therefore, the first components of this analysis are those for which the marginality is maximized, whereas the last components are those on which the specialization is maximized. As such, both the first and the last components are of interest. Thus, the FANTER could be used as a preliminary to the ENFA, because it assesses whether it is of interest to partition the niche inertia into marginality and specialization components.

As for the MADIFA, it is straightforward to show that the jth eigenvalue of the analysis corresponds to the utilization mean of the contributions of the jth component to the Mahalanobis distances between the EUs and the centroid of the availability distribution. The first components are those on which the utilization distribution is the furthest from the availability distribution (i.e., the most marginal), whereas the last components are those on which the used EUs are the most concentrated around the availability mean (the most specialized).

Although the FANTER supposes that the availability distribution is symmetric and unimodal, it does not make the same hypotheses about the niche (contrary to the ENFA and the MADIFA). Thus, this analysis is suitable for the exploration of multimodal niches.

4. Examples

We illustrate here the framework of the GNESFA with the concurrent use of the ENFA, the MADIFA and the FANTER, based on two data sets collected on the chamois (*Rupicapra rupicapra*). For these two datasets, we performed these analyses to develop a conceptual model of the niche-environment system under study. For each analysis, we also tested the significance of the first (and last, for the FANTER) eigenvalue of the analysis using a randomization test. At each of the 500 steps of the randomization process, and for a given analysis, we randomly distributed the chamois locations on the area considered as available to it, and we computed again the eigenvalue of the analysis. We finally compared the observed eigenvalue with the distribution of eigenvalues simulated under the hypothesis of random habitat use, to derive a P-value. We also tested the significance of the marginality value (i.e., m^2), using similar randomization tests (see Basille et al., 2008), to assess the importance of this dimension in the ENFA. Because all the pixels of the raster maps cover the same area, we gave to them an equal availability weight (i.e., $1/N$) in the analyses.

4.1. The chamois population in the Chartreuse mountains

The first data set was collected during censuses of the chamois population carried out in November 1997 in the Chartreuse mountains (French Alps, N. 45.33°, E. 5.80°) by the Departmental association of hunters of Isere. During the census, 239 chamois groups were detected in the study area. More details about the sampling protocol can be found in Michallet (2003). The raster maps of six environmental variables describing the habitat were used to define the ecological space (Table 1). We used the GNESFA to compare the distribution of the locations of the animals (utilization) vs. the distribution of the pixels of the entire area (available) in the ecological space.

Table 1
Variables included in the analyses of habitat selection by 239 chamois groups detected during a census in the Chartreuse mountains (French Alps)

Abbreviation	Description
Slope	Slope
Deciduous	Distance to deciduous woodland
Coniferous	Distance to coniferous woodland
Mixed	Distance to mixed woodland
Open	Distance to open land
Ecotone	Distance to the ecotone between open and forested areas (takes a positive value in open areas and a negative value in closed areas)

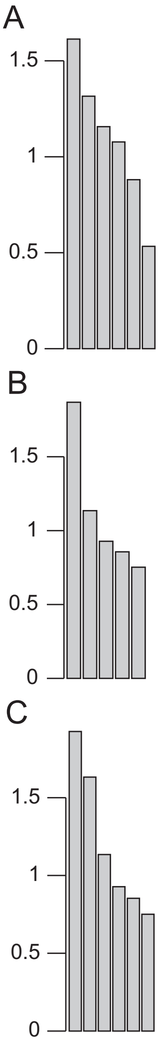

Fig. 3. Barplots showing the eigenvalues of the GNESFA performed to identify correlates between the distribution of chamois detections in the Chartreuse mountain (French Alps) and six environmental variables (listed in Table 1) of the study area: (A) eigenvalues of the FANTER, the first and the last are kept for the analysis, (B) eigenvalues of the ENFA, only the first is kept, (C) eigenvalues of the MADIFA, the first two are kept.

The eigenvalue diagram of the FANTER indicates two patterns in the data, on the first and on the last components of the analysis (Fig. 3A). Indeed, the tests of the first ($\gamma_1 = 1.61$, $P < 0.002$) and of the last eigenvalue ($\gamma_6 = 0.53$, $P < 0.002$) were both significant. A clear "break" is apparent on this diagram before the last

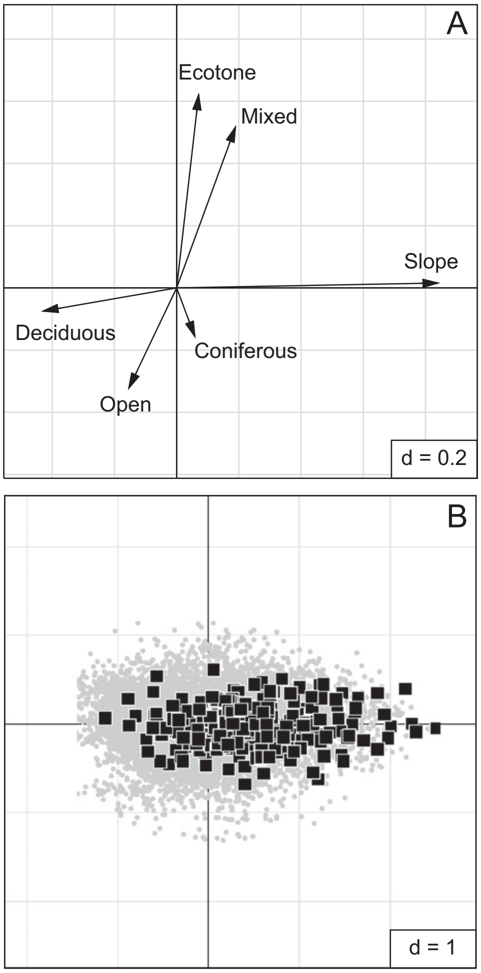

Fig. 4. Results of the FANTER performed to identify correlates between the distribution of chamois detections in the Chartreuse mountain (French Alps) and six environmental variables (listed in Table 1) of the study area. (A) the correlations between the environmental variables and both the first (abscissa) and the last (ordinate) component of the analysis are indicated by arrows. For each variable, the length of an arrow on a given axis gives the value of the correlation between the variable and this component. Grid lines (separated by a distance of 0.2) can be used to measure these correlations on the graph; (B) scatterplot diagram of the cloud of available (grey circle) and used (black squares) points on the first two axes of the MADIFA.

eigenvalue, and a slight break appeared after the first one. The first component was mainly correlated with the slope ($R = 0.84$), while the last was defined by the distance to the ecotone open/forested areas ($R = 0.62$), and to a lesser extent, by the distance to the mixed woodland ($R = 0.51$, Fig. 4A). The factorial map of the niche revealed that the niche was rather marginal on the first axis (Fig. 4B): chamois were rarely located on flat terrain (only 16% of the chamois were detected on slopes lower than 25%, while this habitat type represented 40% of the study area). On the other hand, the specialization of the chamois population was maximum for the distance to the ecotone open/forested areas (25% of the chamois were located at an absolute distance value less than 100 m, while this distance class represented only 13% of the study area) and, to a lesser extent, for the distance to the mixed woodland (40% of the chamois were located between 10 and 150 m from this vegetation type, while this distance class represented only 30% of the study area).

The ENFA confirmed these results. A randomization test of the first eigenvalue of this analysis indicated a significant specialization on at least one component ($S_1 = 1.87$, $P < 0.002$). There was a clear break after the first eigenvalue (Fig. 3B), so that we kept only the first specialization component—and the marginality component as it is a prerequisite of the analysis—for the interpretation. Note that the marginality value in the ENFA was also highly significant ($m^2 = 0.56$, $P < 0.002$). As expected, there was a very strong correlation between the marginality component of the ENFA and the first component of the FANTER ($R = 0.92$), and between the first specialization component of the ENFA and the last component of the FANTER ($R = -0.99$). Thus, in this example, the two analyses highlighted the same patterns. The position of the niche relative to the availability distribution was mainly determined by the slope, while its narrowness was determined by the distance to the ecotone open/forested areas, and to a lesser extent, by the distance to the mixed woodlands.

The test of the first eigenvalue of the MADIFA was also highly significant ($\gamma_1 = 1.93$, $P = 0.008$). The eigenvalue barplot indicated a clear break after the second eigenvalue (Fig. 3C), and we therefore focused our interpretation on the first two components. Actually, the first component of the MADIFA was correlated with both the first specialization component of the ENFA ($R = 0.94$) and the last component of the FANTER ($R = -0.96$). The second component of the MADIFA was strongly correlated with both the marginality component of the ENFA ($R = 0.90$), and the first component of the FANTER ($R = -0.96$). The coordinates of every pixel on the principal components of the MADIFA were combined to compute a reduced-rank environment suitability map (Fig. 5). This is done, for each pixel, by summing its squared coordinates on the first two components of the analysis (Calenge et al., 2008). This map can then be examined to identify the areas where rarely used environmental combinations are found, an additional information which helps to interpret the results.

The three analyses identified similar and consistent patterns in the data, and helped to draw a conceptual model of the niche. Chamois avoided flat terrain and preferred the ecotone between open land and mixed woodland. However, the three analyses were used to deliver different outputs. The FANTER was used as a preliminary analysis to identify the patterns involved in the determination of the niche-environment system, whereas the ENFA distinguished precisely between the patterns determining the position and the spread of the niche. On the contrary, the MADIFA was unable to explicitly disentangle between marginality

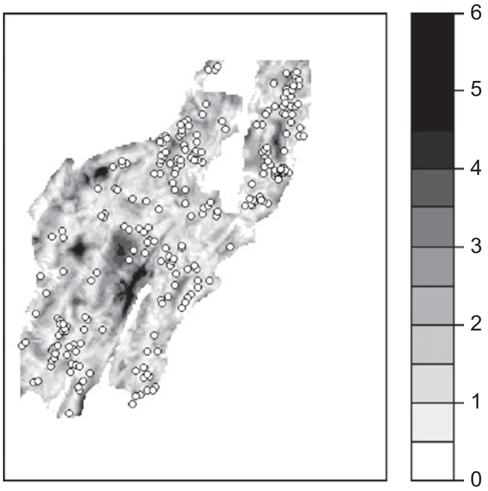

Fig. 5. Environment suitability map for the chamois in the Chartreuse mountain (French Alps), computed by summing the squared coordinates of the pixels on the first two components of the MADIFA (approximate Mahalanobis distance between the pixels and the centroid of the niche). Lighter pixels correspond to suitable areas (low Mahalanobis distance) whereas darker pixels correspond to unsuitable areas (high Mahalanobis distance). The distribution of the chamois detections are also displayed.

and specialization. Rather, it was used to combine these results to compute an environment suitability map. The three analyses returned the same pattern here, but they can highlight dramatically different results in some cases, as demonstrated by the next example.

4.2. *The chamois radio-tracking in the Bauges mountain*

The second dataset describes 56 daily relocations of one chamois in the Bauges mountains (French Alps, N. 45.63°, E. 6.23°). These data were collected from October 1st to November 27th 2003, using a GPS collar. We studied the habitat selection by this individual within its home range (third level of selection according to Johnson, 1980). We therefore computed the home-range limits of this chamois using the minimum convex polygon (Mohr, 1947), and defined the environment by eight environmental variables describing different characteristics potentially important for the chamois, measured for each pixel of a raster map (Table 2). We used the GNESFA to compare the distribution of the relocations of the monitored chamois (utilization) vs. the distribution of the pixels of its home range (available) in the ecological space.

Table 2
Variables included in the analyses of habitat selection by one chamois monitored using a GPS collar in the Bauges mountains (French Alps)

Abbreviation	Description
Elev	Elevation
D.FR	Distance to fallen rocks
D.SeCarS	Distance to meadows made of *Sesleria* and *Carex sempervirens*
D.Trail	Distance to recreational trails
Hydro	Hydrography
Slope	Slope
Sunshine	Sunshine
Visib1000	Visibility computed within a radius of 1000 m

The FANTER indicated a very clear structure driven by the first component (Fig. 6), confirmed by a randomization test of the first eigenvalue ($\gamma_1 = 2.85$, $P = 0.02$). This component contrasted the areas located at high elevations, far from recreational trails, and located on the crests of the mountains (low hydrography), with areas with opposite characteristics. The niche of the individual was bimodal on this component: there was a first mode around the origin, and a second one located on the positive side of this component. Such a pattern can be understood by considering that the shape of the "niche" resulted from both a function giving the probability of selection by the chamois, and by the distribution of the available points. Because the distribution of available points showed a high peak at coordinates around zero, a moderate probability of selection for zero coordinates resulted in a peak at

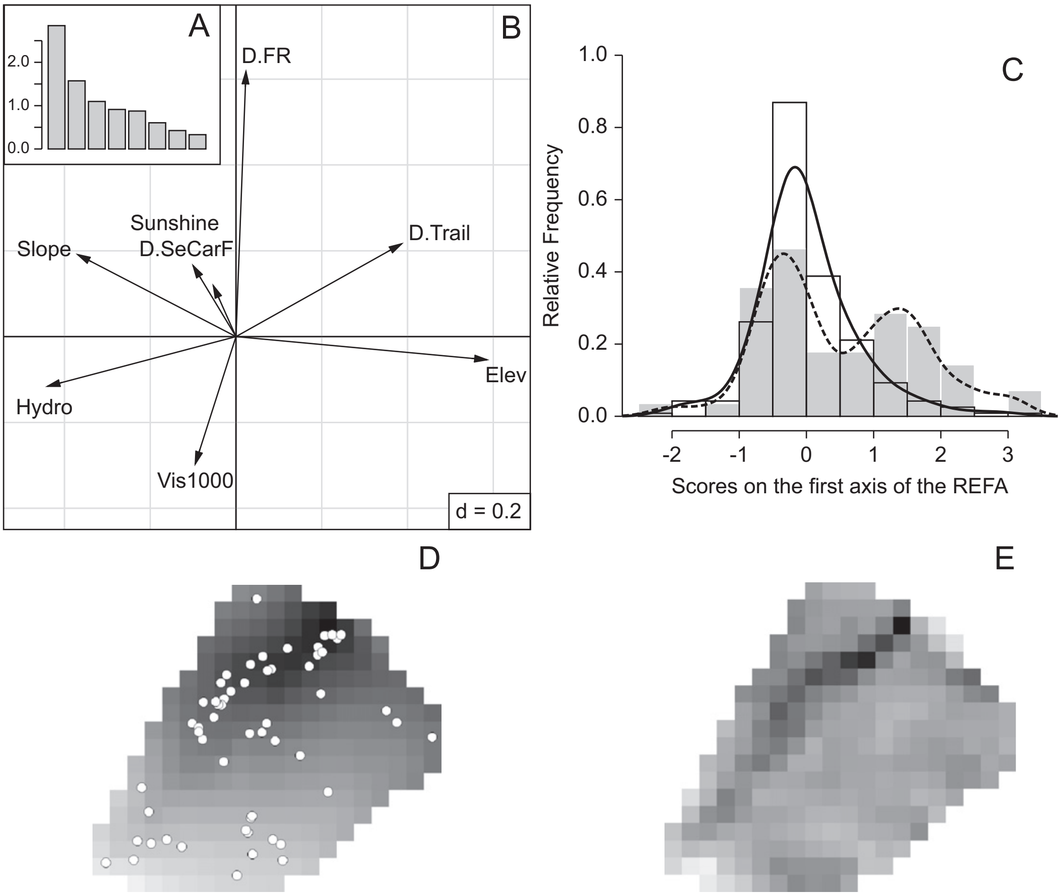

Fig. 6. Results of the FANTER carried out to identify correlates between the distribution of the chamois relocations in the Bauges mountain (French Alps) according to eight environmental variables (listed in Table 2) mapped in its home range. (A) eigenvalue diagram of the analysis. Only the first axis is kept for the analysis; (B) the correlations between the environmental variables and the first (abscissa) and the second (ordinate) component of the analysis. Grid lines (separated by a distance of 0.2) can be used to measure these correlations on the graph; (C) histogram and smoothed density (using a kernel smoothing with smoothing parameter equal to 0.3) of the coordinates of the available points (white bars and continuous curve) and of the used points (grey bars and dashed curve) on the first component of the FANTER; (D) map of the chamois relocations on a map of the elevation in its home range (darker areas are higher); (E) map of the scores of the pixels of the home range on the first component of the FANTER (darker areas are higher).

these coordinates (the proportion of used points was high at zero because the proportion of available points was high at zero). However, the peak on the positive side of this component revealed a strong selection of the mountain crest by this chamois (Fig. 6). In other words, the utilization distribution is a mix between a random use of space by this chamois (same shape as the available distribution) and a search for mountain crest (with a peak on the positive side). Note that the last eigenvalue of the FANTER, on which the specialization was maximized, did not differ significantly from the hypothesis of random habitat use ($\gamma_8 = 0.33$, $P = 0.73$).

The MADIFA did not highlight any particular pattern since the randomization test of the first eigenvalue was not significant ($\gamma_1 = 3.29$, $P = 0.28$). Similarly, the test of the first eigenvalue of the ENFA did not reject the hypothesis of random habitat use ($\gamma_1 = 2.842$, $P = 0.24$). Note that the marginality was significant in the ENFA ($m^2 = 0.44$, $P = 0.01$). Actually, the marginality component of the ENFA was related to the first axis of the FANTER ($R = 0.76$). However, it does not make sense to use the ENFA or the MADIFA in such situations, as they both rely on the hypothesis of unimodal and symmetric niches. In such situations, the FANTER proves to be very useful, allowing one to describe the shape of the niche under study.

5. Discussion

We introduced a new general framework for the analysis of the niche-environment system. Because the GNESFA is by its very nature exploratory, it does not rely on many constraining hypotheses (e.g., no minimum sample size required), which renders it appealing. This framework encompasses three consistent factor analyses. Among these analyses, the FANTER offers a new point of view on the niche-environment system. We also proved that the ENFA and the MADIFA can be viewed as special cases of the GNESFA. Finally we showed that the application of these three methods to biological data may give different outputs, as they are based on different biological points of view of the niche-environment system.

On one hand, the MADIFA corresponds to the point of view of the expert of the studied species: when the biologist has a good prior knowledge of the kind of environment required by the species, an examination of the niche itself is not of prime interest. In this kind of studies, the aim is frequently to determine whether the environment in the study area (the available environment) is similar to the environment the species usually occupies (the reference). This is typically the point of view used for environmental suitability modelling.

On the other hand, the FANTER corresponds to the point of view of the expert of the studied area: when the biologist has a good prior knowledge of the environmental structure in the study area (e.g., correlates between environmental variables), an examination of the availability distribution in the ecological space may not be of prime interest. In this kind of studies, the aim is frequently to identify the patterns of the niche itself, and in what it differs from the study area (classical point of view in habitat selection studies). The patterns identified by the analysis may be due to a particularly strong or low inertia of the niche within the cloud of available points. A strong inertia is likely due to the marginality of the niche, but may also be the result of a multimodal niche. A low inertia indicates that the niche restriction is high on some directions of the ecological space. However, this method will fail to identify the directions of the ecological space where the niche is both marginal and very restricted, as the first characteristic counterbalances the second one (Calenge, pers. obs.). Fortunately, the other analyses belonging to the framework of the GNESFA can be used to detect such cases. The main advantage of the FANTER is that it does not assume the unimodality of the niche, contrary to the other analyses of the framework.

Finally, the ENFA is at the middle point between the FANTER and the MADIFA. Both distributions are used simultaneously as reference and focus distributions. This symmetric role of the two distributions is gained to the detriment of the ecological space, with the loss of one of its dimensions through the marginality vector. The ENFA can be used to distinguish between the position and the narrowness of the distributions relative to each other. Eventually, only the concurrent use of the three methods with graphical displays of the niche within its environment would lead to a consistent knowledge of the system. This statement underlines the interest of the GNESFA as a general framework for the statistical exploration of the ecological niche. In most cases, the two points of view described previously (species and study area) are of interest for the biologist (as in the examples presented in this paper), so that the three complementary analyses may be used concurrently to build a conceptual model of the niche-environment system under study.

The GNESFA is easy to perform with any statistical software, as it relies on a succession of two principal component analyses, which are widely available in most standard statistical packages. In particular, the package adehabitat (Calenge, 2006) for the R software (R Development Core Team, 2006) contains a set of functions allowing the application of the GNESFA (function `gnesfa()`) and several graphical displays of the results. This package also contains numerous functions for managing raster maps, and is especially well-designed for the exploration of the niche-environment system.

We presented here the GNESFA for the exploratory analysis of one single ecological niche. However, radio-tracking studies (involving numerous animals) and multi-species designs are frequent among ecological studies, and most of them aim to identify the common characteristics of the environment affecting the distribution of the organisms under study (whether animals or species). Preliminary results indicated that the framework of the GNESFA can be extended to cover more complex study designs: in particular, canonical OMI analysis (Chessel and Gimaret, 1997) and the eigenanalysis of selection ratios (Calenge and Dufour, 2006) can be reformulated as special cases of the FANTER (Calenge et al., in preparation). These first results also allow for relating the GNESFA to the large family of methods belonging to the duality diagram framework (Escoufier, 1987), a family also containing most factor analysis (principal component analysis, discriminant analysis, etc.). This family has very interesting properties for the exploration of multidimensional spaces, especially in ecology (see Calenge and Dufour, 2006, for a deeper discussion). Further studies are required to clarify the relationships between these analyses and the GNESFA. With the increasing concern of the ecological community for the study of ecological niches subject to climate change, there is an urgent need to have more than one string to our bow, to ensure the reliability of our conclusions. A multi-niche generalization of the GNESFA would allow for a more effective exploration of a species niche within a community in the ecological space, and would allow for the building of habitat suitability maps for several species at once.

Acknowledgements

We warmly thank the Office national de la chasse et de la faune sauvage (ONCFS) for their financial support. We are also grateful to Gaëlle Darmon (University of Lyon), Sonia Saïd (ONCFS) and

Jean-Michel Jullien (ONCFS) for providing the GPS data collected on the chamois in the Bauges mountain, and to the Fédération Départementale des chasseurs de l'Isere for the data on the chamois in the Chartreuse mountain.

Appendix A. Equivalence between the two problems of the GNESFA

In this appendix, we demonstrate that the first problem of the GNESFA

$$\mathbf{Za}_1 = \mathbf{p}_1 \tag{A.1}$$

$$\mathbf{p}_1^t\mathbf{Rp}_1 = 1 \tag{A.2}$$

$$\gamma_1 = \mathbf{p}_1^t\mathbf{Fp}_1 \quad \text{Max} \tag{A.3}$$

is equivalent to the second problem of the GNESFA:

$$\mathbf{g}_1^t\mathbf{g}_1 = 1 \tag{A.4}$$

$$\mathbf{Zg}_1 = \mathbf{y}$$

$$\beta_1 = \frac{\mathbf{y}^t\mathbf{Fy}}{\mathbf{y}^t\mathbf{Ry}} \quad \text{Max} \tag{A.5}$$

We demonstrate this equivalence, with

$$\mathbf{g}_1 = \frac{\mathbf{a}_1}{\sqrt{\mathbf{a}_1^t\mathbf{a}_1}} \tag{A.6}$$

and

$$\mathbf{a} = \frac{\mathbf{g}_1}{\sqrt{\mathbf{g}_1^t\mathbf{Z}^t\mathbf{RZg}_1}} \tag{A.7}$$

And finally, we prove that $\beta_1 = \gamma_1$.

First, we demonstrate that if $\mathbf{a}_1$ is a solution to the first problem, then the use of Eq. (A.6) gives the vector $\mathbf{g}_1$ as a solution to the second problem. Note that the equation (A.6) implies that the condition (A.4) is fulfilled. Moreover,

$$\beta_1 = \frac{\mathbf{g}_1^t\mathbf{ZFZg}_1}{\mathbf{g}_1^t\mathbf{ZRZg}_1} = \frac{\mathbf{a}_1^t\mathbf{Z}^t\mathbf{FZa}_1}{\mathbf{a}_1^t\mathbf{Z}^t\mathbf{RZa}_1} = \mathbf{a}_1^t\mathbf{Z}^t\mathbf{FZa}_1 = \gamma_1$$

It follows that the condition (A.5) is fulfilled. Consequently, if the vector $\mathbf{a}_1$ is a solution to the second problem, the solution $\mathbf{g}_1$ to the first problem can be found using Eq. (A.6).

Now, we prove that if $\mathbf{g}_1$ is a solution to the second problem, then the use of Eq. (A.7) gives the vector $\mathbf{a}_1$ as a solution to the first problem. First note that

$$\mathbf{a}_1^t\mathbf{Z}^t\mathbf{RZa}_1 = \frac{\mathbf{g}_1^t\mathbf{Z}^t\mathbf{RZg}_1}{\mathbf{g}_1^t\mathbf{Z}^t\mathbf{RZg}_1} = 1$$

The condition (A.1) is fulfilled. Moreover,

$$\gamma_1 = \mathbf{a}_1^t\mathbf{Z}^t\mathbf{FZa}_1 = \frac{\mathbf{g}_1^t\mathbf{Z}^t\mathbf{FZg}_1}{\mathbf{g}_1^t\mathbf{Z}^t\mathbf{RZg}_1} = \beta_1$$

and the condition (A.2) is fulfilled.

This completes the demonstration: the two problems are mathematically equivalent, and the relationships between the solutions of the two problems are described in Eqs. (A.7) and (A.6). Furthermore, $\beta_1 = \gamma_1$.

References

Araujo, M., Pearson, R., Thuillier, W., Erhard, M., 2005. Validation of species-climate impact models under climate change. Global Change Biol. 11, 1504–1513.

Barton, D., David, F., 1956. Some notes on ordered random intervals. J. Roy. Stat. Soc. B Met. 18, 79–94.

Basille, M., Calenge, C., Marboutin, E., Andersen, R., Gaillard, J.M., 2008. Assessing habitat selection using multivariate statistics: some refinements of the Ecological-Niche Factor Analysis. Ecol. Modelling 211, 233–240.

Blondel, J., Chessel, D., Frochot, B., 1988. Bird species impoverishment, niche expansion and density inflation in Mediterranean island habitats. Ecology 69, 1899–1917.

Boyce, M., McDonald, L., 1999. Relating populations to habitats using resource selection functions. Trends Ecol. Evol. 14, 268–272.

Browning, D., Beaupré, S., Duncan, L., 2005. Using partitioned Mahalanobis D2(K) to formulate a GIS-based model of timber rattlesnake hibernacula. J. Wildlife Manage. 69, 33–44.

Burnham, K., Anderson, D., 1998. Model Selection and Inference. Springer, Berlin.

Calenge, C., 2006. The package adehabitat for the R software: a tool for the analysis of space and habitat use by animals. Ecol. Model. 197, 516–519, doi:10.1016/j.ecolmodel.2006.03.017.

Calenge, C., Dufour, A., 2006. Eigenanalysis of selection ratios from animal radio-tracking data. Ecology 87, 2349–2355.

Calenge, C., Darmon, G., Basille, M., Loison, A., Jullien, J.M., 2008. The factorial decomposition of the Mahalanobis distances in habitat selection studies. Ecology 89, 555–566.

Calenge, C., Dufour, A., Maillard, D., 2005. K-select analysis: a new method to analyse habitat selection in radio-tracking studies. Ecol. Modelling 186, 143–153.

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

Chessel, D., Gimaret, C., 1997. Niche–ADE-4 modules documentation. Université Lyon 1.

Clark, J., Dunn, J., Smith, K., 1993. A multivariate model of female black bear habitat use for a geographic information system. J. Wildlife Manage. 57, 519–526.

Cleveland, W., 1993. Visualizing Data. Hobart Press, Summit, New Jersey.

Cleveland, W., 1994. The Elements of Graphing Data. Hobart Press, Summit, New Jersey.

Corsi, F., Dupré, E., Boitani, L., 1999. A large scale model of wolf distribution in Italy for conservation planning. Conserv. Biol. 13, 150–159.

Dray, S., Chessel, D., Thioulouse, J., 2003. Co-inertia analysis and the linking of ecological tables. Ecology 84, 3078–3089.

Elith, J., Graham, C., Anderson, R., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R., Huettmann, F., Leathwick, J., Lehmann, A., Li, J., Lohmann, L., Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. Overton, J., Peterson, A., Phillips, S., Richardson, K., Scachetti-Pereira, R., Schapire, R., Soberon, J., Williams, S., Wisz, M., Zimmermann, N., 2006. Novel methods improve prediction of species distributions from occurrence data. Ecography 29, 129–151.

Escoufier, Y., 1987. The duality diagram: a means of better practical applications. In: Legendre, P., Legendre, L. (Eds.), Development in Numerical Ecology. Series G. Springer, Berlin, pp. 139–156.

Farber, O., Kadmon, R., 2003. Assessment of alternative approaches for bioclimatic modeling with special emphasis on the Mahalanobis distance. Ecol. model. 160, 115–130, doi:10.1016/S0304-3800(02)00327-7.

Frontier, S., 1976. Etude de la décroissance des valeurs propres dans une analyse en composantes principales: comparaison avec le modele du baton brisé. J. Exp. Mar. Biol. Ecol. 25, 67–75.

Guisan, A., Zimmermann, N., 2000. Predictive habitat distribution models in ecology. Ecol. Model. 135, 147–186 doi:10.1016/S0304-3800(00)00354-9.

Guisan, A., Lehmann, A., Ferrier, S., Austin, M., McC. Overton, J., Aspinall, R., Hastie, T., 2006. Making better biogeographical predictions of species' distributions. J. Appl. Ecol. 43, 386–392.

Harville, D., 1997. Matrix Algebra From a Statistician's Perspective. Springer, New York.

Hill, M., 1974. Correspondence analysis: a neglected multivariate method. Appl. Stat. 23, 340–354.

Hirzel, A., Arlettaz, R., 2003. Modeling habitat suitability for complex species distributions by environmental-distance geometric mean. Environ. Manage. 32, 614–623.

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

Hutchinson, G., 1957. Concluding remarks. In: Cold Spring Harbour Symposium. vol. 22. Quantitative Biology, pp. 415–427.

Johnson, D., 1980. The comparison of usage and availability measurements for evaluating resource preference. Ecology 61 (1), 65–71.

Knick, S., Dyer, D., 1997. Distribution of black-tailed jackrabbit habitat determined by GIS in southwestern idaho. J. Wildlife Manage. 61, 75–85.

Knick, S., Rotenberry, J., 1998. Limitations to mapping habitat use areas in changing landscapes using the Mahalanobis distance statistic. J. Agric. Biol. Environ. S. 3, 311–322.

Legendre, P., Legendre, L., 1998. Numerical Ecology, second ed. Elsevier Science BV, Amsterdam.

Michallet, D., 2003. Modélisation de la distribution spatiale du chamois de Chartreuse (*Rupicapra rupicapra cartusiana*) par l'analyse des facteurs environnementaux. Master's thesis, Licence Professionnelle de Traitement de l'Information Géographique, IUT de Perpignan.

Mohr, C., 1947. Table of equivalent populations of North American small mammals. Am. Midl. Nat. 37, 223–249.

Morrison, M., Marcot, B., Mannan, R., 1992. Wildlife-Habitat Relationships. Concepts and Applications. The University of Wisconsin Press.

Noy-Meir, I., 1973. Data transformation in ecological ordination. I. some advantages of non-centering. J. Ecol. 61, 329–341.

R Development Core Team, 2006. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0. URL ⟨ http://www.R-project.org⟩.

Reutter, B., Helfer, V., Hirzel, A., Vogel, P., 2003. Modelling habitat-suitability using museum collections: an example with three sympathic Apodemus species from the Alps. J. Biogeogr. 30, 581–590.

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

Seber, G., 1984. Multivariate observations. Wiley Series in Probability and Mathematical Statistics, Wiley, New York.

Soberon, J., Peterson, A., 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. Biodiversity Informatics, 1–10.

Thioulouse, J., Chessel, D., 1992. A method for reciprocal scaling of species tolerance and sample diversity. Ecology 73, 670–680.

Zaniewski, A., Lehmann, A., Overton, J., 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. Ecol. Model. 2002, 261–280 doi:10.1016/S0304-3800(02)00199-0.