N° d'ordre: Année 2004

THESE
présenté devant l'UNIVERSITE CLAUDE BERNARD - LYON I

pour l'obtention du
DIPLOME DE DOCTORAT
(*arrêté du 25 avril 2002*)

Soutenue le 14 décembre 2005

par

Clément CALENGE

Des outils statistiques pour l'analyse des semis de points dans l'espace Ecologique (ANNEXES)

Composition du jury

Directeurs :	Mme Anne-Béatrice DUFOUR M. Daniel MAILLARD
Rapporteurs :	M. Antoine GUISAN M. Patrick DUNCAN
Examinateurs :	M. Christian GAUTIER M. Francis LALOE

UMR CNRS 5558
Laboratoire de Biométrie et Biologie Evolutive
Université Claude Bernard LYON I - Bât G. Mendel
43, boulevard du 11 novembre 1918
69622 Villeurbanne

Sommaire des annexes

Annexe 14 : “How to analyze location data”. Clément Calenge, Paolo Cavallini & Carlotta Canova. Didacticiel de la bibliothèque de fonctions adehabitat disponible sur le site du groupe “Animove", URL: http://www.faunalia.com/animov/howto.php

Annexe 15 : “Package 'adehabitat' ”. Clément Calenge. Manuel d'utilisation de la bibliothèque de fonctions adehabitat

<u>Annexe 1</u>

<u>Article 1 :</u>

“K-select analysis: a new method to analyse habitat selection in radio-tracking studies”

Clément Calenge, Anne-Béatrice Dufour & Daniel Maillard.

Ecological modelling (2005), 186, 143-153

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Ecological Modelling 186 (2005) 143–153

ECOLOGICAL MODELLING

www.elsevier.com/locate/ecolmodel

K-select analysis: a new method to analyse habitat selection in radio-tracking studies

Clément Calenge[a,*], A.B. Dufour[a], D. Maillard[b]

[a] *UMR CNRS 5558, Laboratoire de Biométrie et de Biologie Evolutive, Université Claude Bernard Lyon 1,*
69622 Villeurbanne Cedex, France
[b] *Office national de la chasse et de la faune sauvage, 95 rue Pierre Flourens, F34098 Montpellier Cedex 05, France*

Received 5 August 2004; received in revised form 3 December 2004; accepted 13 December 2004

Abstract

Two kinds of wildlife habitat studies can be distinguished in the literature: hindcasting and forecasting studies. Hindcasting studies aim to emphasize among a large set of habitat variables those that are of interest for the focus species, whereas forecasting studies are intended to predict habitat selection according to a small number of habitat variables for a given area. We provide here a new analytical tool which relies on the concept of ecological niche, the K-select analysis, for hindcasting studies of habitat selection by animals using radio-tracking data. Each habitat variable defines one dimension in the ecological space. For each animal, the difference between the vector of average available habitat conditions and the vector of average used conditions defines the marginality vector. Its size is proportional to the importance of habitat selection, and its direction indicates which variables are selected. By performing a non-centered principal component analysis of the table containing the coordinates of the marginality vectors of each animal (row) on the habitat variables (column), the K-select analysis returns a linear combination of habitat variables for which the average marginality is greatest. It is a synthesis of variables which contributes the most to the habitat selection. As with principal component analysis, the biological significance of the factorial axes is deduced from the loading of variables. An example is provided: habitat selection by wild boar is studied in a Mediterranean habitat using the K-select analysis. The numerous advantages of the analysis (a large number of variables that can be included, individual variability in habitat selection taken into account, a lack of too strict underlying hypotheses) make it a powerful approach in radio-tracking studies designed to identify habitat variables that are selected by animals.

Keywords: Eigenanalysis; Functional responses; Hindcasting studies; Marginality; Radio-tracking

1. Introduction

The concept of habitat has been defined as the resources and conditions present in an area that produce

---

* Corresponding author. Tel.: +33 4 72 43 27 57;
fax: +33 4 72 43 13 88.
E-mail address: calenge@biomserv.univ-lyon1.fr (C. Calenge).

0304-3800/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2004.12.005

occupancy – including survival and reproduction – by a given organism (Hall et al., 1997). Although the habitat concept is by essence multivariate, most studies carried out on habitat selection by animals consider only one single categorical habitat variable, e.g. using the analysis selection ratios (Manly et al., 2002) or compositional analysis (Aebischer et al., 1993). Focusing on only one variable in habitat selection studies has therefore been criticized by many authors (Erickson et al., 1998; Morrison, 2001). The recent increase in availability of geographic information system (GIS) facilitates the study of selection of multiple habitat variables by animals (Knick and Dyer, 1997). Statistical methodology has also been improved and many methods are now available to study habitat selection (Manly et al., 2002). These methods generally compare the habitat use and availability to highlight the most required habitats. Thomas and Taylor (1990) have distinguished three kinds of study designs in the literature, according to the level at which habitat availability and use are measured. With design I, both availability and use are measured at the population level (the individual animals are not identified; Erickson et al., 1998). With design II, habitat availability is again measured at the population level, but the use is measured for each animal (some individual animals belonging to the population are identified). With design III, both the availability and use of habitat are measured for each identified animal. Designs II and III generally involve the monitoring of animals using radio-tracking (Aebischer et al., 1993), and are the most frequent designs in the literature dealing with habitat selection—about 75% of the studies (Schooley, 1994).

Moreover, Morrison et al. (1992) also make a distinction between hindcasting and forecasting studies of habitat selection. Hindcasting identifies key environmental variables that account for observed variation in species variables, whereas forecasting involves an explicit attempt to predict species habitat use. Forecasting studies are generally carried out with the help of the resource selection functions methodology (Manly et al., 2002; Boyce and McDonald, 1999). Resource selection functions, defined like any functions proportional to the probability of use by an organism, are generally fitted in a generalized linear model framework. Habitat suitability maps may then be derived from such models. Actually, the fit of these models supposes “that the modeller knows the limiting factors that influence the distribution and abundance of the study organism” (Boyce and McDonald, 1999). This point is also stressed by Guisan and Zimmermann (2000): “One of the most difficult tasks is to decide which explanatory variables, or combination of variables, should enter the model”. It is sometimes possible to select these factors from the results of the ecological literature (Burnham and Anderson, 1998), but more often, the intention of habitat selection studies is precisely to identify such factors. Therefore, hindcasting should necessarily precede forecasting. We stress that a given statistical approach will not necessarily be as efficient for both objectives.

Several techniques are generally used for the exploration of habitat selection with observational data as entries. Among them, generalized linear models are often used “by subjectively and iteratively searching data for patterns and significance” (Burnham and Anderson, 1998, p. 17). It is well known, however, that such “data-dredging” may lead to strongly overfitted models with poor biological meaning and an unknown validity. Automatic variable selection methods such as stepwise regression are also commonly applied, but this strategy is considered to be “fishing expedition” by many statisticians and biologists (Johnson, 1981; Guisan et al., 2002; Hirzel et al., 2002). The exploration of data may thus quickly become an obscure operation in modeling techniques.

On the other hand, eigenanalyses have already proved their efficiency in revealing the main characteristics of multivariate data (Escoufier, 1987). They are an extension of the principal component analysis (PCA). An eigenanalysis is characterized by a triplet of matrices: (i) a table to be analysed, (ii) a matrix of weights for the columns of this table, and (iii) a matrix of weights for its rows. The weighted PCA of the table assigns scores to rows and columns so that the sum of the squared row scores (inertia) is maximized and the successive axes returned by the analysis are orthogonal (uncorrelated). The table to be analysed and the weight matrices vary depending upon the method used. Simple geometrical methods such as PCA and correspondence analysis, or more complex methods such as discriminant analysis and canonical correlation analysis, are eigenanalyses (e.g. see Dolédec et al., 2000). Such methods have frequently been used with success in ecological studies (Manly, 1994) because their graphical approaches allow conclusions to be drawn based mainly on a visual interpretation of the data, which is a very attractive feature for the biologists. Therefore, eigenanalyses have a central place in the early steps of exploratory data analysis.

We propose here a new multivariate approach intended to emphasize the limiting factors in habitat use in designs II and III hindcasting studies: the K-select analysis. We derive a geometric method for exploratory studies relying on the concept of ecological niche (Hutchinson, 1957). We especially focus on marginality (Hausser, 1995; Dolédec et al., 2000; Hirzel et al., 2002), a criterion that measures the squared Euclidean distance between the average habitat conditions used by an organism and the average habitat conditions available to it. An eigenanalysis of the marginality vectors is performed to summarize the habitat selection common to all animals. This method is illustrated in an application on radio-tracking data collected on the wild boar in a Mediterranean habitat. And finally we present the numerous qualities of the K-select analysis.

2. A Niche-based approach to analyse habitat selection

2.1. The marginality

We assume that the studied area includes a set of discrete resource units (RU), on which J habitat variables are measured (Manly et al., 2002). These RUs may be, for example, the pixels of a raster map. A sample of K animals has been captured and monitored by radio-tracking so that the number of relocations of the animal k in a RU gives an estimate of its use. Thus we have information on available and used resource units, which is termed “sampling protocol A” by Manly et al. (2002). We also assume that the relocations are temporally independent for all animals, an assumption necessary for the randomization tests developed below. Thus, the position of a relocation does not depend on the position of the previous one (Swihart and Slade, 1985). The question is how to highlight the factors affecting the use of space by animals (hindcasting studies).

The niche of a species was defined by Hutchinson (1957) as the hypervolume in the multidimensional space of environmental variables where this species can maintain a viable population. In a mathematical way, the niche of a species can be viewed as a multivariate probability density function which gives the density of probability of the species presence according to the position in the ecological space. If the niche can be assumed to be multivariate normal, the mean vector of this distribution is the optimum location for the species and defines the point where the probability density of use is the highest. The squared distance of this optimum from the point located at the average of available habitat conditions on the study area is called “marginality” (Hausser, 1995; Hirzel et al., 2002) and measures the strength of habitat selection, i.e. the mean difference between habitat use and availability. The derivation of methods based on the marginality allowing the analysis of habitat selection using designs II and III implies a formal mathematical definition of this criterion.

First, consider the case of only one animal k. The $I_k \times J$ matrix $\mathbf{X}_k$, contains the measurements of the J habitat variables in the set of the I_k RUs available to the animal k. Centering and standardizing this matrix by column gives the table $\mathbf{X}_k^*$. This table defines a cloud of I_k points in a J-dimensional space (Fig. 1a). Since the table $\mathbf{X}_k^*$ is centered, the origin O_k of this scatterplot is located at the average available habitat conditions. Let $\mathbf{v}_k$ be a I_k-dimensional vector that contains the number of observations of the animal k in each RU. Let $\mathbf{f}_k = \mathbf{v}_k / \sum_{i=1}^{I_k} v_{ki}$, where v_{ki} is the number of observations of the animal k in the ith RU; $\mathbf{f}_k$ contains the relative frequency of use of the RUs by the organism. Thus, the vector $\mathbf{m}_k = \mathbf{f}_k^t \mathbf{X}_k^*$ is the marginality vector for this animal, and its squared norm, the marginality of the animal, measures the squared distance between the average available habitat conditions (point O_k in Fig. 1a) and the habitat conditions used by the animal (point G_k in Fig. 1a).

A randomization test can be performed to determine whether habitat selection is significant (Manly, 1997). This test is achieved by considering the equiprobability of the random allocation of the $\sum_{i=1}^{I_k} v_{ki}$ relocations in the RUs available to the animal k and by re-computing the marginality for randomized data sets. The observed marginality is then compared to the randomized values of marginality to determine whether selection is significant for this set of variables. Although rather simple, these geometrical foundations may easily be extended to analyse other kinds of designs.

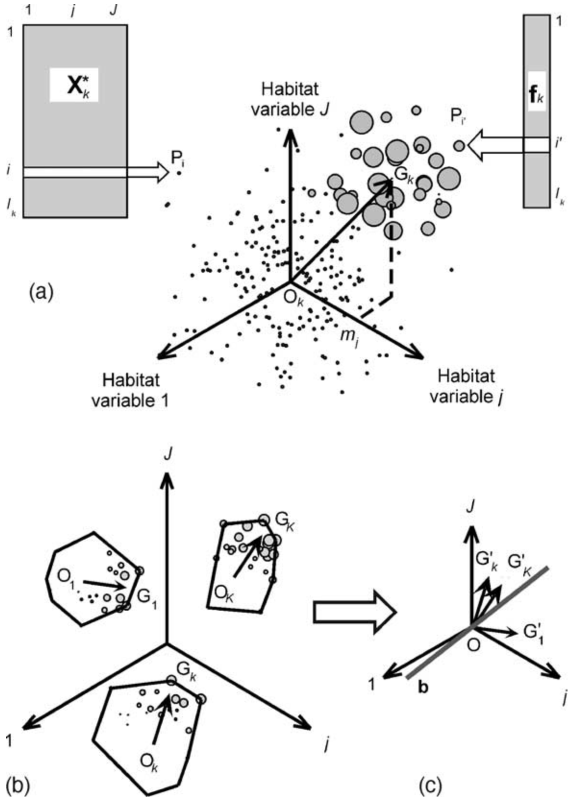


Fig. 1. Analysis of the marginality. (a) Case of one animal: the table $\mathbf{X}_k^*$, centered and standardized by columns, contains the values of J habitat variables on the I_k resource units available to the animal k. Each row of $\mathbf{X}_k^*$ defines a point in the J-dimensional space of habitat variables. The origin O_k of the space is located at the average of available habitat conditions. The vector $\mathbf{f}_k$ contains the relative frequencies of use of each unit by the animal. The diameter of the circles is proportional to these frequencies. The average m_j of the habitat variable j is weighted by the relative frequencies of use. The m_js are the coordinates of the point G_k, which is located at the average used habitat conditions. The vector O_kG_k is the marginality vector of the animal k. (b) Case of K animals (design III): For each animal k, the average available habitat conditions define a point O_k and the average used conditions define a point G_k. The vector O_kG_k is the marginality vector for animal k. (c) The K-select analysis proceeds in two steps: first, a translation is applied to each vector O_kG_k, so that they all have a common origin O (the origin of space); second, an eigenanalysis is performed on the table of coordinates of the translated vectors $OG_k{}'$ on habitat variables, so that the mean marginality projected on the first axis **b** is maximized.

2.2. The K-select analysis: a new method to analyse designs II and III data

In design III, the availability of habitat may vary from animal to animal, whereas it is the same for all animals in design II. Thus, a method developed for design III can also be used in design II studies, by setting all availabilities equal to the same value for each habitat variable. We therefore use here the above-described geometrical approach to build a multivariate analysis of the marginality for design III data, keeping in mind that the analysis can be extended to design II studies. We consider here a radio-tracking study of K animals. The RUs available to a given animal may be determined in different ways, e.g. all pixels located within a specified distance from an animal relocation. The global table $\mathbf{X}$ contains the value of the J habitat variables for all K animals:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \mathrm{M} \\ \mathbf{X}_K \end{bmatrix}$$

This table has J columns and I rows, with $I = \sum_{k=1}^{K} I_k$. Each table $\mathbf{X}_k$ is then transformed in $\mathbf{X}_k^{\bullet}$ so that each column of the global table $\mathbf{X}^*$:

$$\mathbf{X}^* = \begin{bmatrix} \mathbf{X}_1^{\bullet} \\ \mathbf{X}_2^{\bullet} \\ \mathrm{M} \\ \mathbf{X}_K^{\bullet} \end{bmatrix}$$

is centered and has unit variance. For the animal k, the $1 \times J$ vector $\mathbf{a}_k$,

$$\mathbf{a}_k = \frac{1}{I_k} \mathbf{1}_{I_k}^{\mathrm{t}} \mathbf{X}_k^{\bullet}$$

contains the coordinates of the mean of habitat conditions available to the animal k (Fig. 1b). Let $\mathbf{f}_k$ be the I_k-dimensional vector that contains the proportion of the relocations of animal k numbered in each RU. Further, let $\mathbf{u}_k = \mathbf{f}_k^{\mathrm{T}} \mathbf{X}_k^{\bullet}$. The vector $1 \times J$ $\mathbf{u}_k$ contains the coordinates of the mean of habitat conditions used by animal k. Therefore, the vector $\mathbf{m}_k^{\bullet} = \mathbf{u}_k - \mathbf{a}_k$ is the marginality vector of animal k (see Fig. 1c). These vectors may be concatenated by row to form the $K \times J$ table $\mathbf{M}$:

$$\mathbf{M} = \begin{bmatrix} \mathbf{m}_1^{\bullet} \\ \mathbf{m}_2^{\bullet} \\ \mathrm{M} \\ \mathbf{m}_K^{\bullet} \end{bmatrix}$$

The table **M** contains the coordinates of the K marginality vectors on the J habitat variables. Let d_k be the proportion of all relocations collected for the animal k:

$$d_k = \frac{I_k}{\sum_{k=1}^{K} I_k}$$

Let $\mathbf{D}_K = \mathrm{Diag}(d_1, \ldots, d_k, \ldots, d_K)$ be the weight matrix associated with the rows of the matrix **M**. The animals thus have a weight proportionate to the number of their relocations. Note that under some circumstances, the biologists may prefer to give uniform weights to all animals rather than weights proportionate to the number of their relocations. In this case, the animal weights are given by $d_k = 1/K$ instead of the above formula. The three matrices ($\mathbf{M}$, $\mathbf{I}_J$, $\mathbf{D}_K$) define a statistical triplet that can be analysed using an eigenanalysis. That is, the $K \times J$ table **M** is to be analysed, the $J \times J$ identity matrix $\mathbf{I}_J$ defines the weights associated with the columns (square matrix with 1s on the diagonal and 0s elsewhere), and the $K \times K$ matrix $\mathbf{D}_K$ defines the weights associated with the rows. The total inertia of this analysis is equal to

$$\mathrm{Tr}(\mathbf{M}^{\mathrm{T}}\mathbf{D}_K\mathbf{M}\mathbf{I}_J) = \sum_{K=1}^{K} d_k ||\mathbf{m}^{\bullet}_k||^2,$$

i.e. equal to the mean marginality of the monitored animals, weighted by the number of relocations of each animal. Therefore, an eigenanalysis of the triplet ($\mathbf{M}$, $\mathbf{I}_J$, $\mathbf{D}_K$) gives the linear combinations of habitat variables that maximize the mean marginality of the monitored animals on the first axes. On the resulting graphical display, the average of used habitat conditions is as far as possible from the average of available habitat conditions. This eigenanalysis is strictly equivalent to the non-centered PCA of the table **M**, with the uniform weighting for all columns and weights proportional to the number of animal relocations for the rows. This analysis indicates the similarity of the habitat selection among several animals. When all the animals have the same habitat preferences, their marginality vectors are oriented in the same direction, and the first axes of the analysis explain a large part of the total inertia. Conversely, if the habitat selection differs from animal to animal, the component of the marginality explained by the first axes is not much larger than the inertia explained by the last.

An idea of the pertinence of this analysis can also be obtained using randomization tests. If we assume that the relocations are temporally independent, then under the null hypothesis of random habitat use, all the possible allocations of the $\sum_{i=1}^{I_k} v_{ki}$ relocations of animal k in the RUs available to it are equiprobable. “Random habitat use” marginality vectors can then be computed from this randomized data set, and a table **M** can be computed. This procedure is repeated several times and with each step, a K-select analysis is carried out and the first eigenvalue is stored. Finally, the first eigenvalue of the observed data set is compared to the distribution of the first eigenvalues of the randomized data sets. The tests presented in the case of design I data can also be applied to each individual using the Bonferroni correction (Bland and Altman, 1995; test of the marginality and test of the effects of variables on the marginality of each animal).

All these equations are given for the case of continuous variables, yet it is essential to note that this approach still holds for dummy (Tenenhaus and Young, 1985), fuzzy coded (Chevenet et al., 1994) or a mixture of quantitative and dummy variables (Hill and Smith, 1976). This analysis may therefore be used to analyse all kinds of habitat variables (categorical and/or continuous).

3. Application to wild boar (*Sus scrofa* L.) data

3.1. Description of the data

The procedure described above was applied to one data set concerning radio-tagged wild boars monitored in Puéchabon, France (Maillard, 1996). The study area (4500 ha) is located in the south of France and the vegetation is typical of the Mediterranean habitat (garrigue dominated by *Quercus ilex* L.). The study area was mapped according to nine continuous variables: elevation, slope, sunshine, distance from water, recreational trails and human areas, and density of herbaceous, shrubby and tree vegetation (Fig. 2). Two main topographical structures may be identified. The Hérault gorges pass through our study area and affect nearly all of our ecological variables. Indeed, this valley is characterized by low elevation, steep slopes, low sunshine, less dense herbaceous cover and is generally far from recreational trails. The Puéchabon plateau, delimited

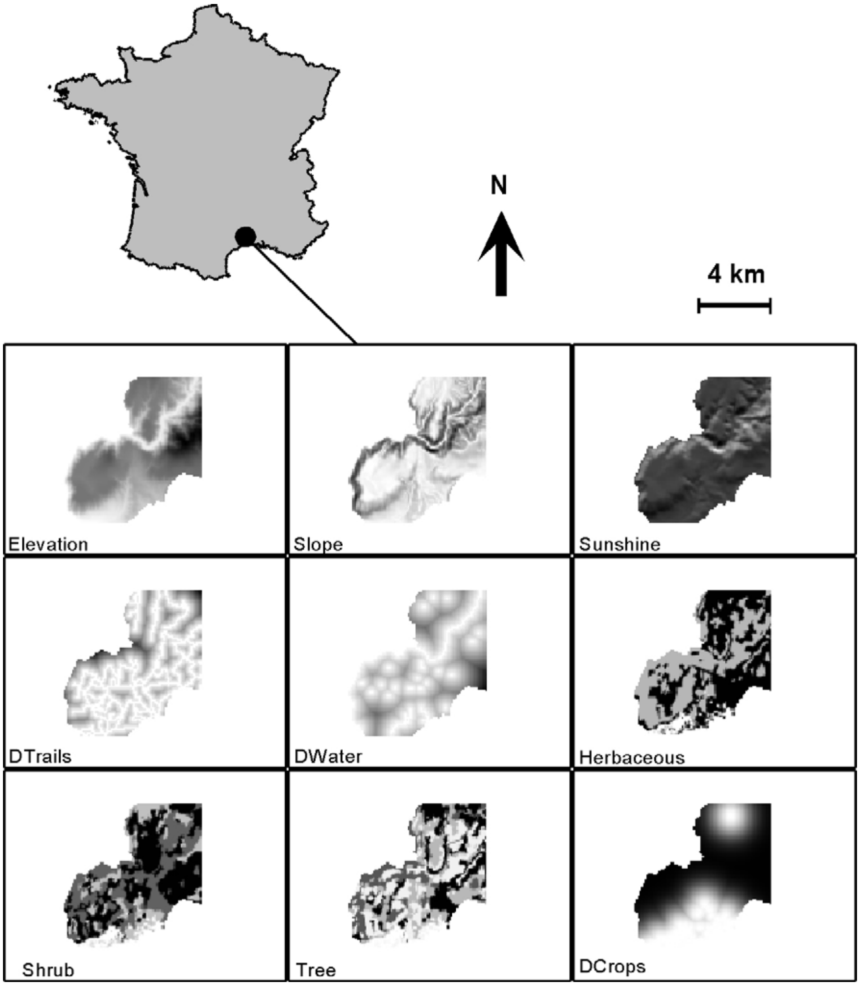


Fig. 2. Location and maps of the study area for wild boar radio-tracking at Puéchabon, south of France. The studied variables are elevation, slope, sunshine, distance from recreational trails (Dtrails), distance from water (Dwater), density of vegetal cover (herbaceous, shrub and tree cover) and distance from crops (Dcrops).

by the gorges, is the core of our study area. Nine trap sites baited with maize were placed on this plateau in summer 1992 and 1993. Eleven wild boars were captured and fitted with radio collars. Animals were then relocated daily at their resting sites (Maillard, 1996). We focus here on diurnal resting-site selection in summer (July–August). All of the pixels on the raster map located within 500 m from the relocations of a given animal were considered to be available to it, each pixel covering an area of one hectare. The habitat availability thus varied from one animal to another. We then applied the K-select analysis to this data. All analyses were carried out using the R software (Ihaka and Gentleman, 1996).

3.2. Results

Habitat use appeared significantly non-random for nearly all animals, as indicated by the randomization tests carried out on the marginality vectors (Table 1a). However, it is more difficult to learn from results of the tests carried out for each variable and each animal (Table 1b). These tests are rarely significant, even when the α-level is shifted to 0.1. Indeed, the

Table 1
Results of the randomization tests of habitat selection by the wild boar at Puéchabon (design III data; see text)

	WB1	WB2	WB3	WB4	WB5	WB6	WB7	WB8	WB9	WB10	WB11
(a) Tests of the marginality (Bonferroni α level = 0.05/11 = 0.0045)											
Marginality	1.058	2.600	3.057	1.499	0.452	2.280	0.634	1.089	1.145	2.412	2.452
# relocations	50	52	54	82	51	34	48	57	23	51	51
P-value	0.004	<0.0001	<0.0001	<0.0001	0.155	<0.0001	0.073	0.008	0.007	0.0003	0.0003
(b) Selection of habitat variables by each animal (Bonferroni α level = 0.05/99 = 0.0005, two-tailed test)											
Elevation	0.33	−0.93[*]	−0.94[*]	0.37	−0.26	−0.55	−0.33	0.19	−0.27	−0.93[*]	−0.86[*]
Slope	−0.69[*]	0.39	0.26	0.05	0.02	0.48	−0.16	−0.17	−0.35	0.37	0.50
Sunshine	0.16	−0.45	−0.43	0.09	−0.09	−0.65[*]	0.22	−0.16	−0.32	−0.56	−0.53
Trails	−0.19	0.75[a]	1.17[*]	−0.14	0.21	0.72	0.27	0.07	−0.15	0.46	0.96[*]
Water	0.05	−0.64	−0.61	0.77[a]	−0.06	−0.51	−0.36	0.29	−0.30	−0.60	−0.35
Herbaceous	−0.35	−0.27	−0.30	0.34	−0.47	−0.45	−0.12	0.83[a]	−0.51	−0.26	−0.19
Shrub	−0.06	0.11	−0.14	0.78[*]	−0.02	−0.07	−0.13	0.2	0.44	0.03	0.22
Tree	0.53	0.56	0.20	0.12	−0.27	0.27	−0.29	0.39	−0.53	0.67	0.16
Crops	0.03	0.03	0.07	−0.14	0.18	0.53	−0.37	0.18	0.08	0.03	0.07

These tests are based on 10,000 randomization steps (NS: not significant). (a) significance of the marginality for each animal. (b) Coordinates of marginality vectors on habitat variable (i.e. the differences [mean used – mean available] for each wild boar and each variable). Using the mathematical notation presented in the text, this table corresponds to the table $\mathbf{M}^{\mathrm{T}}$ (transpose of $\mathbf{M}$). The column names correspond to the animals' ID code.

[a] Significant at the 10% level.

[*] Significant at the 5% level.

biggest drawback of the Bonferroni inequality is that it becomes even more conservative when a large number of tests are carried out (Bland and Altman, 1995, Faraway, 2002). In addition, habitat variables are correlated among each other. It is therefore rather tricky to emphasize the important underlying factors from such tests. Finally, variation exists in habitat selection from animal to animal. For example, WB2 and WB3 exhibit similar habitat preferences, with selection of low elevations and selection of areas far from trails, whereas WB1 selects the steep slopes. Because of this variation of preferences, it is difficult to summarize the patterns of habitat selection from only these results.

This individual variability in habitat selection should not be ignored (Mysterud and Ims, 1998) and the K-select analysis may be used to summarize the patterns of habitat choice. The first eigenvalue λ_1 of the analysis is larger than what is expected under the random habitat use hypothesis ($\lambda_1 = 1.203$, $P < 0.0001$). In other words, the K-select analysis is pertinent, and there is at least one group of several animals all displaying some similarity in habitat preferences. Since this method is a PCA-like analysis, the results should be analysed in the same manner. The barplot of eigenvalues indicates the amounts of marginality explained by each factorial axis (Fig. 3a). The amount of explained marginality decreases regularly after the second axis, indicating that the following axes account for "random noise". We thus analysed the marginality on the first two axes, keeping in mind that the first axis explained a larger part of the marginality than the second.

The biological interpretation of the axes is done according to the variable loadings (Fig. 3b). The first axis was explained by an opposition between the Hérault gorges (close to water and far from recreational trails, with low elevation, higher slopes and low sunshine) and the Puéchabon Plateau (further from water and closer to trails, with high elevation, low slopes and high sunshine). The second axis is mainly explained by the density of vegetation cover (herbaceous, shrub and tree cover). Three main groups of animals may be identified (Fig. 3c). A first group strongly selected the Hérault gorges for their resting site (WB11, WB10, WB2, WB6 and WB3). Animals of the second group preferred the plateau (WB4 and WB8), but also strongly selected a dense vegetation cover for their resting sites. The third group of animals included the boars (WB1, WB9, WB7 and WB5) with a weaker or even non-significant marginality (Table 1a).

A graphical display of the niche of a given animal is obtained by projecting the cloud of available pixels on the first two factorial axes (Fig. 4). This is done by

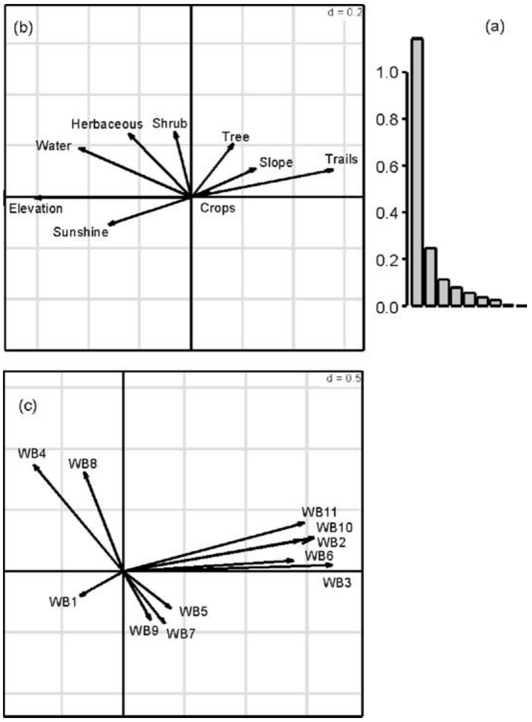


Fig. 3. Results of the K-select analysis carried out to highlight habitat selection by 11 wild boars on nine habitat variables: (a) bar chart of the eigenvalues, measuring the mean marginality explained by each factorial axis; (b) variable loadings on the first two factorial axes; (c) projection of the marginality vectors of all animals on the first factorial plane. All marginality vectors are recentered such that habitat availability is the same for all animals (common origin to all vectors).

projecting the matrix $\mathbf{X}^*$ onto the axes of the analysis. The origin of this graph is the barycenter of the pixels available to all animals, since $\mathbf{X}^*$ is centered by column. We then add black squares with areas proportional to the number of relocations counted in the pixels. A star connecting all used pixels identifies the animal niche in the space of available resources. Finally, the contour polygon of available pixels (light grey in Fig. 4) and of used pixels (dark grey) completes the representation of the niche of the animal. The position of the available pixels relative to the origin gives an idea about the accessibility of the main characteristics highlighted by the analysis to a given animal.

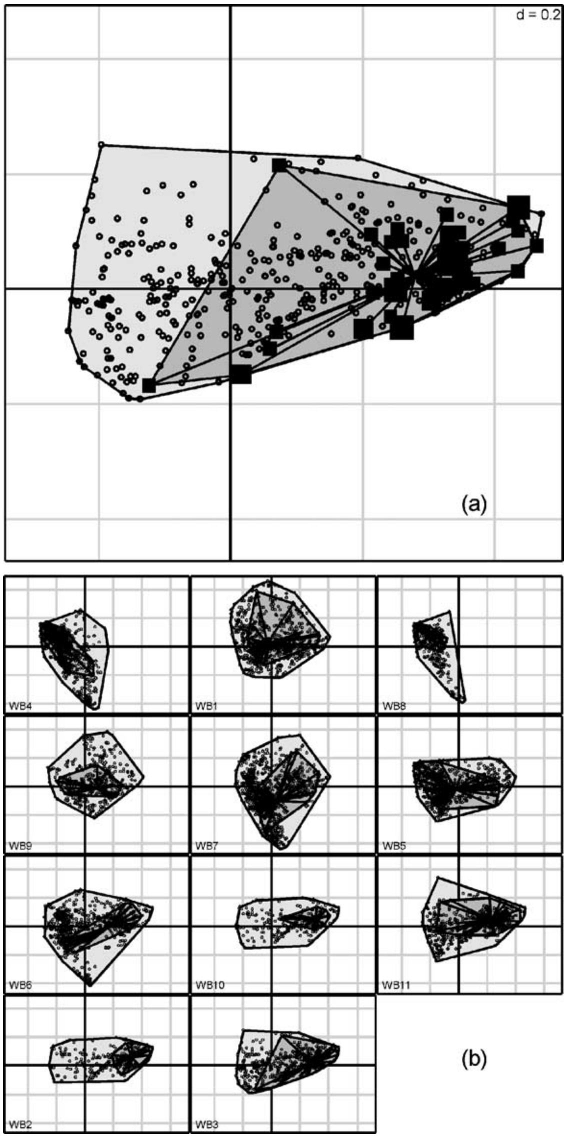


Fig. 4. Representation of the niche of the wild boars monitored at Puéchabon on the first two factorial axes. (a) Detail for wild boar #3, (b) The same graphical display for all monitored animals. Note that the graphs are ordered according to the coordinates of the corresponding animals on the first factorial axis (negative values are at the upper left hand corner and positive values at the lower right hand corner).

The difference in habitat selection between the three groups may thus be explained by the difference in habitat availability for the animals (“functional responses”, Mysterud and Ims, 1998). Most animals select the gorges, a favorable place for resting. The absence of recreational trails and the steep slopes encountered in this area ensure the absence of human disturbance, while the lower sunshine occurring in the gorges is certainly a key factor in their selection, especially during summer, in the warm Mediterranean environment. The gorges were very accessible for this group of animals. Moreover, the second group (WB4 and WB8) had greater limitations in access to these areas (Fig. 4). Their home range was indeed located on the Plateau. Thus, they selected the densest vegetation areas for their resting sites, probably to limit human disturbance.

We stress that these conclusions are only valid for the scale of the study (Johnson, 1980). Availability as defined in this example can also measure habitat use in studies carried out on a larger spatial scale. For example, all the pixels on the raster map located within 500 m from a relocation could be used to characterize the selection of the home range within the study area. If the habitat selection was studied on a larger scale, these pixels could have been considered as used, and availabilities could have been set equal to the composition of the study area. In this case, the K-select analysis might have helped to identify the most important factors in the selection of home range within the study area. The K-select analysis can therefore be used for data analysis in multi-scale analysis of habitat selection.

4. Discussion

A new method is developed to point out the habitat selected by a species in designs II and III studies. The K-select analysis led us to identify the main features of the habitat selected by the wild boar for its resting site in a Mediterranean environment. We thus detected a functional response in habitat selection, i.e. a variation of the habitat selection according to the habitat availability which we used to build a typology of the animals. This analysis is therefore suitable when the objective is to define one or several groups of animals that select the same habitat characteristics.

Wildlife habitat studies generally use other kinds of eigenanalyses, such as principal component analysis or multiple correspondence analysis. This is viewed as a preliminary step to reduce the number of variables before further predictive analyses (e.g. Brown and Batzli, 1984 on the squirrel; Abaigar et al., 1994 on the wild boar). This strategy has been severely criticized, however, because there is no indication that the first factorial axes are those of interest to the focus species, even if they account for most of the spatial variability in the study area. As noted by Johnson (1981), "there is no reason to assume that a principal component necessarily relates to the animal or its needs. The animal could be responding to one of the variables that is a minor component of all that were measured".

The K-select analysis focuses on the differences between habitat use and availability. This approach consists of a PCA performed on a non-centered table. As stated by Noy-Meir (1973), "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". Thus, by considering the average available habitat conditions as the origin of ecological space, an eigenanalysis of the table containing the relative position of the average used habitat conditions finds the linear combination of habitat variables for the habitat selection which is the largest. These synthesis variables are of strong ecological interest for the stated objectives, especially for the identification of the habitat variables selected by animals.

This analysis is neither restricted by the number of habitat variables nor by the number of animals, which is another very attractive characteristic of the method. A strong correlation among habitat variables has no negative effect on the results. Even a small number of animals allows for the description of the relationships between habitat use and availability. It can limit, however, the biological conclusions drawn from the analysis and therefore influence the inferential power of the technique.

The K-select analysis focuses on only one parameter of the niche, the marginality. This criterion measures the squared distance between the optimum for an animal and the average available habitat conditions. Actually, another parameter which could be of interest to biologists is the tolerance (Dolédec et al., 2000; Hirzel et al., 2002), i.e. the variance of habitat use around this optimum on habitat variables. It is worth noting two points here, however: (i) performing the randomization tests on marginality vectors is equivalent to testing the importance of the marginality relative to the tolerance for each animal: the test is significant when the marginality is large relative to the tolerance; (ii) the K-select analysis focuses on the "between-animal" variability, not on the "within-animal" variability. Indeed, in habitat selection studies relying on radio-tracking data, the sampling unit of interest is often the animal and rarely the relocation (Aebischer et al., 1993; Otis and White, 1999).

The application of the K-select analysis only requires that the relocations of the animals are independent and representative of the actual habitat use of the animals. The sensitivity of the analysis to temporal autocorrelation between the relocations merits further investigation. Temporal autocorrelation occurs when the time interval between two relocations is short, because, in that case, the position of a relocation depends on the position of the previous one (Swihart and Slade, 1985). The K-select analysis in itself is not affected by autocorrelation, as this analysis is mainly exploratory. However, when this hypothesis is violated, the randomization tests of the marginality vectors are not valid since all permutations of the frequencies of use in the available resource units are not equiprobable. In other words, when autocorrelation is present in the data, one works with trajectories of the animals, and no longer with a random sample of relocations. This point needs to be taken into account in the randomization tests. For example, Martin (2004) recommends generating random walks with the same properties as the observed trajectories (same distance between successive relocations). The simulated data can then be used as an estimation of random habitat use in the randomization tests, instead of a random sample of independent points.

Here, we have provided a powerful tool to identify, among a set of habitat variables, those that are of interest in resource selection studies based on designs II and III data. With the development of new technologies such as GIS or GPS, biologists who formerly had meager data sets now have to deal with too much data. Among all the variables potentially useful to describe the habitat of a species, only a limited number should be used for prediction of habitat use (Boyce and McDonald, 1999). For this reason, the first axes of a K-select analysis could be included in a predictive model of habitat use. We stress that this perspective merits further investigation.

5. Software availability

This analysis, as well as other computational tools, is available in the “adehabitat” package for R software (Ihaka and Gentleman, 1996), and can be downloaded for free at the URL http://cran.r-project.org/.

Acknowledgements

We are grateful to the Office national de la chasse et de la faune sauvage for their financial support. We warmly thank Daniel Chessel (University of Lyon, France), for his help in the development of the K-select analysis, and Jean-Michel Gaillard (University of Lyon, France), for his invaluable comments on earlier drafts of this paper.

References

Abaigar, T., Del Barrio, G., Vericad, J.R., 1994. Habitat preference of wild boar (*Sus scrofa* L., 1758) in a Mediterranean environment. Indirect evaluation by signs. Mammalia 58, 201–210.

Aebischer, N.J., Robertson, P.A., Kenward, R.E., 1993. Compositional analysis of habitat use from animal radio-tracking data. Ecology 74, 1313–1325.

Bland, J.M., Altman, D.G., 1995. Multiple significance tests: the Bonferroni method. Br. Med. J 310, 170.

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

Brown, B.W., Batzli, G.O., 1989. Habitat selection by fox and gray squirrels: a multivariate analysis. J. Wildl. Manage. 48, 616–621.

Burnham, K.P., Anderson, D.R., 1998. Model selection and inference. In: A Practical Information-theoretic Approach. Springer Verlag, Berlin, Germany.

Chevenet, F., Doledec, S., Chessel, D., 1994. A fuzzy coding approach for the analysis of long-term ecological data. Freshwater Biol. 31, 295–309.

Dolédec, S., Chessel, D., Gimaret-Carpentier, C., 2000. Niche separation in community analysis: a new method. Ecology 81, 2914–2927.

Erickson, W.P., Mc Donald, T.L., Skinner, R., 1998. Habitat selection using GIS data: a case study. J. Agric. Biol. Environ. Stat. 3, 280–295.

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

Faraway, J.J., 2002. Practical Regression and ANOVA using R. Department of Statistics, University of Michigan, USA.

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Model 135, 147–186.

Guisan, A., Edwards, T.C., Hastie, T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecol. Model 157, 89–100.

Hall, L.S., Krausman, P.R., Morrison, M.L., 1997. The habitat concept and a plea for standard terminology. Wildl. Soc. Bull. 25, 173–182.

Hausser, J., 1995. Säugetiere der Schweiz. Mammifères de la Suisse. In: Mammiferi della Svizzera. Birkhäuser Verlag, Berlin, Germany.

Hill, M.O., Smith, A.J.E., 1976. Principal component analysis of taxonomic data with multi-state discrete character. Taxon 25, 249–255.

Hirzel, A.H., 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.E., 1957. Concluding remarks. Cold Spring Harbor Symp. Quantitat. Biol. 22, 415–427.

Ihaka, R., Gentleman, R., 1996. R: A language for data analysis and graphics. J. Comput. Graphic. Stat. 5, 299–314.

Johnson, D.H., 1981. The use and misuse of statistics in wildlife habitat studies. In: Capen, D.E. (Ed.), The use of Multivariate Statistics in Studies of Wildlife Habitat, U.S. Fish and Wildlife Service. USDA Forest Service, pp. 11–19.

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

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

Maillard, D., 1996. Occupation et utilisation de la garrigue et du vignoble méditerranéens par le Sanglier. Thesis. University of Aix-Marseille III, Marseille, France.

Manly, B.F.J., 1994. Multivariate Statistical Methods. A primer, 2nd ed. Chapman & Hall, London, UK.

Manly, B.F.J., 1997. Randomization, Bootstrap and Monte Carlo Methods in Biology, 2nd ed. Chapman & Hall, London, UK.

Manly, B.F.J., McDonald, L.L., Thomas, D.L., McDonald, T.L., Erickson, W.P., 2002. Resource Selection by Animals: Statistical Design and Analysis for Field Studies, 2nd ed. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Martin, J., 2004. Importance des structures spatiales et des contraintes comportementales dans les études de sélection de l'habitat. Msc Thesis, Université Claude Bernard Lyon 1, Lyon.

Morrison, M.L., Marcot, B.G., Mannan, R.W., 1992. Wildlife–Habitat Relationships. Concepts and Applications. The University of Wisconsin Press, Madison, WI.

Morrison, M.L., 2001. A proposed research emphasis to overcome the limits of wildlife-habitat relationship studies. J. Wildl. Manage. 65, 613–623.

Mysterud, A., Ims, R.A., 1998. Functional responses in habitat use: availability influences relative use in trade-off situations. Ecology 79, 1435–1441.

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

Otis, D.L., White, G.C., 1999. Autocorrelation of location estimates and the analysis of radiotracking data. J. Wildl. Manage. 63, 1039–1044.

Schooley, R.L., 1994. Annual variation in habitat selection: patterns concealed by pooled data. J. Wildl. Manage. 58, 367–374.

Swihart, R.K., Slade, N.A., 1985. Testing for independence of observations in animal movements. Ecology 66, 1176–1184.

Tenenhaus, M., Young, F.W., 1985. An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data. Psychometrika 50, 91–119.

Thomas, D.L., Taylor, E.J., 1990. Study designs and tests for comparing resource use and availability. J. Wildl. Manage. 54, 322–330.

Annexe 2

Article 2 :

“Factorial analysis of selection ratios from animal radio-tracking data”

Clément Calenge & Anne-Béatrice Dufour.

Ecology, soumis

Running head: FACTORIAL ANALYSIS OF SELECTION RATIOS

FACTORIAL ANALYSIS OF SELECTION RATIOS FROM ANIMAL RADIO-TRACKING DATA

C. CALENGE,[1,2] AND A.B. DUFOUR [1]

[1] *UMR CNRS 5558, Laboratoire de biométrie, Université Claude Bernard Lyon 1, 69622 Villeurbanne Cedex, France.*

---

[2] Corresponding author.

E-mail: calenge@biomserv.univ-lyon1.fr

Abstract. The development of methods to analyse habitat selection when resources are defined by several categories (e.g. vegetation types) is a topical issue in radio-tracking studies. The White and Garrott statistic, an extension of the widely used Neu et al.'s test, can be used to determine whether habitat selection is significant. As well, the Manly's selection ratio, a particularly useful measure of resource selectivity by resource users, allows detection of the most strongly selected habitat types. However, when both the number of animals and habitat types are large, the biologist often has to deal with an excessive large number of measures. In this paper we present a new method, the factorial analysis of selection ratios, which generalizes these two common methods within the framework of eigenanalyses. The total inertia of this analysis is equal to the White and Garrott statistic, which is maximized on the first factorial axes. The factorial analysis of selection ratios is therefore optimal in habitat selection studies. Although we mainly consider the case where the habitat availability is the same for all animals (design II), this analysis is also extended to the case where the habitat availability varies from one animal to another (design III). An application of this method is provided using radio-tracking data collected on 17 squirrels in five habitat types. The results indicate variability in habitat selection, with two groups of animals displaying two patterns of preference. This difference between the two groups is explained by the patch structure of the study area. Because this method is mainly exploratory, and therefore does not rely on any distributional assumption, we recommend its use in studies of habitat selection.

Keywords: chi-square, design II, design III, factorial analysis, eigenanalyses, selection ratios, radio-tracking, habitat type

INTRODUCTION

It is a major point of concern for both ecologists and managers to investigate which habitats are selected by animals, to draw solid conclusions on the ecology of species and thereby increase the efficiency of management policies. Habitat selection studies usually compare the use of habitats by animals with their availability. For a long time, such studies have focused on only one single categorical variable defining several habitat types, e.g. vegetation types (Hall et al. 1997). The recent increase in availability of Geographic Information Systems has rendered easier the inclusion of a larger number of habitat variables in such studies (e.g. slope, distance from water points, etc.); as well, statistical methodology have been developed which allows the analysis of these more complex schemes (Manly et al. 2002). However, the development of methods for the analysis of habitat selection when resources are defined by several categories is still a topical issue and new approaches are still regularly proposed (Aebischer et al. 1993, Conner and Plowman 2001, Manly et al. 2002).

The analysis of habitat selection generally consists of the comparison of habitat use and habitat availability. The methods allowing for the study of habitat selection may be divided into three categories, according to the type of study design on which they rely (Thomas and Taylor 1990): with design I, both habitat availability and habitat use are measured at the population level (the individual animals are not identified, Erickson et al. 1998); with design II, animals are identified and used habitat is measured for each, but the habitat availability is defined at the population level, i.e. it is the same for all animals; and with design III, both the availability and use of habitat are measured for each animal. Design II and III studies generally involve the monitoring of animals using radio-tracking (Aebischer et al. 1993). Whatever the study design involved, two issues are usually raised in habitat selection studies: (i) whether selection is significant with regard to the habitat types considered and (ii) which habitat types are selected by animals.

Among the numerous methods available for question (i), the most commonly used is the Neu et al. ’s (1974) test (Schooley 1994). It basically consists of a chi-square test comparing the use of habitat types with their availability. However, the assumption of independence between sampling units is likely to be violated in design II and III studies (Otis and White 1999). Two relocations of one single animal are indeed more dependent than two relocations of two animals, and to cope with this problem, White and Garrott (1990) proposed a modification of this test for design II and III studies. It consists of computing one chi-square per animal, and then summing these statistics to obtain a global measure of the significance of habitat selection. This approach was also recommended by Manly et al. (2002).

Conversely, several methods have been proposed to emphasize habitat types selected by animals, and hence answer question (ii) (Johnson 1980, Aebischer et al. 1993, Cherry 1996, Conner and Plowman 2001, Manly et al. 2002). Among them, the selection ratios proposed by Manly et al. (1972) have become a classical approach. These measures are computed for each habitat type and for each animal as the ratio of the used proportion to the available proportion. These criteria are the best estimates of the relative probabilities of selection of habitat types (Manly et al. 2002), i.e. the probabilities multiplied by an unknown constant. However, if I habitat types and J animals are considered, the biologist has to deal with $I \times J$ selection ratios, which becomes intractable when the number of both animals and habitat types increases. Manly et al. (2002) proposed to work instead on the average of selection ratios per habitat type. The implicit hypothesis underlying this approach is that the selection ratios of animals are drawn from a normal (or at least unimodal) distribution. In other words, one supposes that on average, all animals select the same habitat types. In reality, numerous factors, internal (e.g. age or sex, Aebischer et al. 1993) as well as external (e.g. habitat availability, Mysterud and Ims 1998), may affect habitat selection and fends to increase its variability. That is, some habitats may be strongly selected by some animals and weakly selected by others. Therefore, averaging selection ratios should not be done without first investigating variability.

The eigenanalyses may be useful to achieve this objective (Greenacre 1984, Escoufier 1987). This method family groups all analyses that can be considered as extensions of Principal Component Analysis (PCA). An eigenanalysis is characterized by a triplet of matrices: (i) a table to be analyzed; (ii) a matrix of weights for the columns of this table; and (iii) a matrix of weights for its rows. The weighted PCA of the table assigns scores to its rows and its columns, so that the inertia of the row scores is maximized and the successive axes returned by the analysis are two by two orthogonal (uncorrelated). The table to be analyzed and the weight matrices vary with the method used (Greenacre 1984, Doledec et al. 2000). Common methods such as discriminant analysis, canonical correlation analysis, correspondence analysis or multidimensional scaling belong to this group.

We propose here a new eigenanalysis for studies of habitat selection with design II and III data to be applied when habitats are defined by several categories. This approach derives from both White and Garrott (1990)'s test and the analysis of selection ratios. This method is suitable for the exploration of habitat selection, and allows identification of animal groups which select their habitat in a common way. An example of application is shown using the squirrel data set provided by Aebischer et al. (1993), and illustrates the richness of this approach.

MATHEMATICAL MATERIAL

We first focus on the analysis of design II data. Let I be the number of habitat types considered. A random sample of J animals is obtained from the population and monitored using radio-tracking. The relocations by the j^{th} animal are sampled to estimate the proportions used of different types of habitat categories, and the area considered as available to the population of animals is also sampled or censused to estimate the proportions of area in each of the several habitat categories.

Let u_{ij}, for $1 \leq j \leq J$ and $1 \leq i \leq I$, be the number of relocations by animal j in habitat type i. Moreover, let

$$u_{i+} = \sum_{j=1}^{J} u_{ij} \qquad \text{(total for habitat } i\text{)}$$

$$u_{+j} = \sum_{i=1}^{I} u_{ij} \qquad \text{(total for animal } j\text{)}$$

and

$$u_{++} = \sum_{i=1}^{I} \sum_{j=1}^{J} u_{ij} \qquad \text{(total number of relocations of all animals)}$$

Let p_i be the proportion of available resource units in habitat type i. Manly et al. (2002) proposed the use of selection ratios w_{ij} to measure habitat selection in design II studies, when resources are defined by several categories:

$$w_{ij} = \frac{u_{ij}}{u_{+j} p_i}$$

Under the hypothesis that the animal j uses habitat type i randomly, this ratio is on average equal to 1. The larger this value is, then the stronger the selection for this habitat is. Manly et al. (2002) proved that this ratio is the best estimation of the relative probability of selection of habitat type i by animal j.

Moreover, several tests have been proposed to test the overall habitat selection by animals with this kind of design, the most common being the chi-squared test proposed by White and Garrott (1990), which is carried out in two steps. First, for each animal, the statistic

$$\chi_j^2 = \sum_{i=1}^{I} \frac{(u_{ij} - p_i u_{+j})^2}{p_i u_{+j}}$$

is computed for each animal *j*. Under the null hypothesis of random habitat use, this statistic is drawn from a chi-squared distribution with I-1 degrees of freedom.

The second step consists of summing these statistics for all animals:

$$\chi^2_{WG} = \sum_{j=1}^{J}\sum_{i=1}^{I}\frac{(u_{ij} - p_i u_{+j})^2}{p_i u_{+j}}$$

Note that this statistic may be written:

$$\chi^2_{WG} = \sum_{j=1}^{J}\sum_{i=1}^{I} p_i u_{+j}\left(\frac{u_{ij} - p_i u_{+j}}{p_i u_{+j}}\right)^2 = \sum_{j=1}^{J}\sum_{i=1}^{I} p_i u_{+j}\left(\frac{u_{ij}}{p_i u_{+j}} - 1\right)^2$$

equivalent to

$$\chi^2_{WG} = \sum_{i=1}^{I}\sum_{j=1}^{J} p_i u_{+j}(w_{ij} - 1)^2 \qquad \text{(Eq. 1)}$$

Under the random hypothesis of random habitat use by all animals, this statistic follows a Chi-square distribution with $J\times(I\text{-}1)$ degrees of freedom. This approach was also proposed by Manly et al. (2002).

Furthermore, the hypothesis of identical habitat selection for all animals is tested by a standard chi-square on the contingency table containing the u_{ij}'s (Manly et al. 2002), i.e.

$$\chi^2_c = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\left(u_{ij} - u_{i+}.u_{+j}\right)^2}{u_{i+}.u_{+j}}$$

This statistic is to be compared with the standard chi-square distribution with (I-1) × (J-1) degrees of freedom.

THE FACTORIAL ANALYSIS OF SELECTION RATIOS

Let the matrix **W** (I rows × J columns) contain the centered selection ratios:

$$\mathbf{W} = \left[w_{ij} - 1\right]_{i=1\ldots I,\, j=1\ldots J}$$

Under the hypothesis of random habitat use, (w_{ij} – 1) is equal to 0. Thus, the origin of the column space, i.e. a J-dimensional vector of 0, corresponds to a hypothetical animal using all habitat types randomly. Similarly, the origin of the row space corresponds to a hypothetical habitat type used in proportion to its availability by all animals.

Let the $J \times J$ matrix $\mathbf{D} = Diag(u_{+1}...u_{+j}...u_{+J})$ be the weight matrix associated with the columns of **W**. Let the $I \times I$ matrix $\mathbf{P} = Diag(p_1...p_i...p_I)$ be the weight matrix associated with the rows of **W**.

The factorial analysis of selection ratios is the eigenanalysis of the statistical triplet (**W**, **P**, **D**). The weighted and non-centered principal component analysis of the **W** table assigns scores that maximize the square deviation between animals and this hypothetical animal using space randomly. It also assigns scores to the habitat types that maximize the square distance between habitat types and a habitat type used in proportion to its availability.

The use of **P** as column weight ascertains that the more common habitat types will have more weight than habitats scarcely encountered. Similarly, the use of **D** as row weight implies that more importance is given to animals with more relocations. The inertia of this analysis is equal to:

$$\text{Inertia} = Tr({}^{t}\mathbf{WDWP})$$ where ${}^{t}\mathbf{W}$ means the transpose of **W**.

This is equivalent to

$$\text{Inertia} = \sum_{i=1}^{I}\sum_{j=1}^{J} p_i u_{+j}(w_{ij}-1)^2 = \chi^2_{WG}$$ where χ^2_{WG} is the statistic computed in equation (1).

Thus, the inertia of this analysis is equal to the White and Garrott (1990)’s statistic. This statistic is therefore maximized on the first axes, or rather the first axes are those for which habitat selection is the highest. This analysis is optimal to study habitat selection with design II and may also be extended to design III in a straightforward way (See Appendix).

APPLICATION

The procedure described above was applied to the squirrel data set provided by Aebischer et al. (1993) which describes habitat use and availability for 17 radio-tracked squirrels. Thirty relocations per animal were collected at a rate of three radio locations per day over a 10-d period. Five habitats types were defined by the authors: Young, Thuja, Larch, Mature and Open. Each type occurred as patches within well-defined boundaries. The limits of the study area were defined as the boundaries of all habitat patches containing at least one relocation, plus those which overlapped any home range or were surrounded by such patches. We consider here the selection of the relocations of the animals within the study area (second order habitat selection, Johnson 1980).

The White and Garrott's Chi-square test for habitat selection is highly significant (χ^2_{WG} = 1563.9, df = 68, P < 0.001), indicating a strong habitat selection by animals. Furthermore, the selection was not identical for all animals (χ^2_c = 321.7, df = 64, P < 0.001). The factorial analysis of selection ratios has been carried out to explain this heterogeneity.

The analysis produces two factors explaining a large part of the White and Garrott's chi-square statistic (90.6%), and the percentage of information explained is nearly equal for the two axes (49.6% for the first axis, and 41% for the second). The Open and Thuja habitat types were rarely used by all animals (Fig. 1), whereas the three remaining habitats were frequently used. However, there is a large variability in habitat selection among animals. In fact, we can identify two strategies: (i) some animals primarily used the Larch habitat type, and used the Mature and Thuja less frequently; (ii) others preferred the Larch relative to the Mature and Young habitat type. This is confirmed by the study of the selection ratios; the highest selection ratios are indeed obtained from the Young habitat for the animals of the first group, and from the Larch for the animals of the second group (Table 1). The Mature habitat was second in rank for 13 out of the 17 animals.

This variability in habitat selection highlighted by the analysis may be explained by the patch structure of the habitat and the home-range size of the squirrels. The study area includes only one patch of Young and one patch of Larch habitat type and the two patches are 70 meters apart. The home-range size of the monitored squirrels, computed using the minimum convex polygon (Mohr 1947), covers on average of 1.75 ha (S.E. = 0.23). A circle with the same area would have a 75 meters radius. As the home range is a life history trait which is characteristic of the species in a given environment (Benhamou 1998), the two patches of searched habitat types (Young and Larch) were too far apart to allow the squirrels to use both of them. Even if the Larch and the Young habitat types are searched for by the animals, because they are not contiguous, animals must choose one strategy of habitat selection. Therefore, the patch structure may explain why two groups appear on the first factorial plane of the analysis. One group of animals established its home range in the Larch habitat type and also used the neighbouring habitat types (Mature), though less intensively. Conversely, another group selected the Young habitat and also used the neighbouring types.

The importance that patchiness of habitat types has on selection has already been reported in habitat selection studies and has been related to the concept of scale (Otis 1998). As noted by previous authors, several studies of habitat selection carried out on a given area at several scales will not necessarily give the same results (Johnson 1980, Levin 1992). Therefore, the next step of the analysis would be the description of the habitat selection to a finer degree, by restricting what is considered as available to the animals. For example, we may investigate the selection of the relocations within the home range (Aebischer et al. 1993). As the availability would then vary from one animal to another, this step could be carried out using the factorial analysis of selection ratios extended to design III (see Appendix).

DISCUSSION

We developed here a new graphical tool to investigate habitat selection when habitat is defined by several categories. This analysis, optimal in habitat selection studies, within the framework of eigenanalyses, includes the selection ratios and White and Garrott's measure of habitat selection.

A habitat type used less than its availability is characterized by a selection ratio ranging from 0 to 1. Conversely, a habitat type used more than its availability is characterized by a selection ratio ranging from 1 to infinity. The factorial analysis of selection ratios therefore gives more importance to habitats strongly selected than to habitats scarcely used. This is an important quality of the method, as the selection ratios should not be interpreted as evidence of absolute preference or avoidance (Johnson 1980). The apparent animal's avoidance of an habitat type leads almost invariably to the apparent preference of another type, because of the following constraint: the sum of all used proportions is equal to 1 (Aebischer et al. 1993). Due to this dependence upon habitat types, the selection ratio of a given type only has a meaning in comparison to the others. The higher a given selection ratio is then the stronger the selection is for the corresponding type.

The factorial analysis of selection ratios presents several advantages over the simple averaging of selection ratios. Indeed, averaging these measures over the animals to obtain a single ordering of the habitat types may even obscure important information on the individual variability in habitat selection because this ordering may cover a wide range of situations. For example, an ordering of the habitat type "A greater than B" may indicate that the habitats A and B are present in the home ranges of all animals, but also that as a whole A is preferred to B. However, another explanation may be that some animals used only the habitat A, other animals only the habitat B, and that there are more present in A than in B. This is the case in this paper. In the first explanation, the preferred habitat is present in all home-ranges, which may suggest that this habitat is necessary for the animals. Conversely, the second explanation shows that the animal may survive without the habitat A. This distinction can be of importance when setting a wildlife management or conservation policy.

The same problem may be encountered with all methods that give a single ordering of habitats, without first considering the effects of other factors on habitat selection. This point has led to the development of modelling methods which test the differences between groups of animals (Aebischer et al. 1993, Manly et al. 2002). For example, Heisey (1985) showed the close connection which exists between the log-linear modelling and the selection ratios. Using such models, it is possible to test whether pre-specified variables actually have an effect on the habitat selection. However, the number of factors that can be included in the analyses is often restricted by the number of animals at hand (Aebischer et al. 1993). In addition, all these analyses imply that the biologist knows *a priori* which variables potentially affect habitat selection, which is not always the case, as evidenced by our example. We could not guess *a priori* that the effect of the patch structure in the study area was so important.

On the other hand, in the eigenanalysis framework, the analyst takes “the best image” of the data according to specified criteria, and then searches *a posteriori* the variables that may explain the largest aspect of variability in the data. This approach has already proven its efficiency in ecological studies (Hill 1974, Ter Braak 1985, Blondel et al. 1988, Thioulouse and Chessel 1992). When all the animals select the same habitat types, then use of the first axis of the analysis can explain the greater part of the White and Garrott’s statistic, and thus a ranking of the habitats according to their preferences can be read on this axis. On the other hand, when there is some variability in habitat selection among monitored animals, then the method returns several axes. A graphical display of these results allows for establishing a typology of animals according to the similarity in their habitat selection. From the examination of the data, the mechanisms underlying the habitat use could then be hypothesized.

One of the most positive properties of eigenanalyses is its wide applicability. Indeed, the factorial analysis of selection ratios does not rely on any distributional assumption, and does not make any hypothesis about the data. In addition, it is not mathematically limited by the number of habitat types defined or the number of animals. It is not natural to say that the number of individuals is not important, because it necessarily affects the inferential power of the technique. Another main quality of the eigenanalyses is that this methodology can be used to analyse a large diversity of designs (Escoufier 1987, Doledec et al. 2000, Dray et al. 2003, Thioulouse et al. 2004).

In this paper, we considered only one table of selection ratios (with animals in columns and habitat types in rows), but the analysis of selection ratios may as well be developed for more complex schemes. For example, a given set of animals could have been monitored during several seasons, and the biologist might be interested in the changes of habitat selection between these periods. This aspect could be treated with the help of the K-tables analyses, a family of methods extending the principle of eigenanalyses to the analysis of several tables matched by rows and/or by columns (Thioulouse et al. 2004). However, this point surely needs further research. Because of its numerous qualities, the factorial analysis of selection ratios may find its place in the exploratory analysis of habitat selection by animals monitored by radio-tracking in cases where the resources are defined into several categories.

SOFTWARE AVAILABILITY

This analysis may be computed using the ade4 package for R. The extension to design III can be carried out using the R package adehabitat. The R software can be freely downloaded from the internet at URL: http://cran.r-project.org/

ACKNOWLEDGEMENTS

We are grateful to the "Office national de la chasse et de la faune sauvage" for their financial support. We also thank Robb Ogden (Wall Street Institute, Lyon) for correcting the English.

REFERENCES

Aebischer, N. J., P. A. Robertson, and R. E. Kenward. 1993. Compositional analysis of habitat use from animal radio-tracking data. Ecology **74**:1313-1325.

Benhamou, S. 1998. Le domaine vital des mammifères terrestres. Revue d'écologie - La Terre et la Vie **53**:309-334.

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

Cherry, S. 1996. A comparison of confidence interval methods for habitat use-availability studies. Journal of Wildlife Management **60**:653-658.

Conner, L. M., and B. W. Plowman. 2001. Using Euclidean distances to assess nonrandom habitat selection. Pages 275-290 *in* J. Marzluf, editor. Radio tracking and animal populations. Academic Press, San Diego, California, USA.

Doledec, S., D. Chessel, and C. Gimaret-Carpentier. 2000. Niche separation in community analysis: a new method. Ecology **81**:2914-2927.

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

Erickson, W. P., T. L. McDonald, and R. Skinner. 1998. Habitat selection using GIS data: a case study. Journal of Agricultural, Biological, and Environmental Statistics **3**:296-310.

Escoufier, Y. 1987. The duality diagram: a means of better practical applications. Pages 139-156 *in* L. Legendre, editor. Development in numerical ecology. Springer Verlag, Berlin.

Greenacre, M. J. 1984. Theory and applications of correspondence analysis. Academic Press, London, UK.

Hall, L. S., P. R. Krausman, and M. L. Morrison. 1997. The habitat concept and a plea for standard terminology. Wildlife Society Bulletin **25**:173-182.

Heisey, D. M. 1985. Analysing selection experiments with log-linear models. Ecology **66**:1744-1748.

Hill, M. O. 1974. Correspondence analysis: a neglected multivariate method. Applied Statistics **23**:340-354.

Johnson, D. H. 1980. The comparison of usage and availability measurements for evaluating resource preference. Ecology **61**:65-71.

Levin, S. A. 1992. The problem of pattern and scale in Ecology. Ecology **73**:1943-1967.

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.

Manly, B. F. J., P. Miller, and L. Cook. 1972. Analysis of a selective predation experiment. American Naturalist **106**:719-736.

Mohr, C. O. 1947. Table of equivalent populations of North American small mammals. American Midland Naturalist **37**:223-249.

Mysterud, A., and R. A. Ims. 1998. Functional responses in habitat use: availability influences relative use in trade-off situations. Ecology **79**:1435-1441.

Neu, C. W., C. R. Byers, and J. M. Peek. 1974. A technique for analysis of utilization-availability data. Journal of Wildlife Management **38**:541-545.

Otis, D. L. 1998. Analysis of the influence of spatial pattern in habitat selection studies. Journal of Agricultural, Biological, and Environmental Statistics **3**:254-267.

Otis, D. L., and G. C. White. 1999. Autocorrelation of location estimates and the analysis of radiotracking data. Journal of Wildlife Management **63**:1039-1044.

Schooley, R. L. 1994. Annual variation in habitat selection : patterns concealed by pooled data. Journal of Wildlife Management **58**:367-374.

Ter Braak, C. J. F. 1985. Correspondence analysis of incidence and abundance data : properties in terms of a unimodal reponse model. Biometrics **41**:859-873.

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

Thioulouse, J., M. Simier, and D. Chessel. 2004. Simultaneous analysis of a sequence of paired ecological tables. Ecology **85**:272-283.

Thomas, D. L., and E. J. Taylor. 1990. Study designs and tests for comparing resource use and availability. Journal of Wildlife Management **54**:322-330.

White, G. C., and R. A. Garrott. 1990. Analysis of wildlife radio-tracking data. Academic press, London, UK.

APPENDIX

We demonstrate here how the factorial analysis of selection ratios may be extended to the analysis of design III data.

For design II, this analysis consists in the eigenanalysis of the triplet (**W**, **P**, **D**). It is straightforward to show that this method is equivalent to the eigenanalysis of the triplet (**S**, $\mathbf{I}_{JJ}$, **D**), where $\mathbf{I}_{JJ}$ is the $J\times J$ identity matrix (square matrix with 1 on the diagonal and 0 elsewhere), and **S** (I rows × J columns) is the following matrix**:**

$$\mathbf{S} = \left[\sqrt{p_i}\,(w_{ij} - 1)\right]_{i=1\ldots I,\, j=1\ldots J}$$

In designs III, availability is defined for each animal. Let $p_{i/j}$ be the proportion of habitat type i available to animal j so that

$$\sum_{i=1}^{I} p_{i/j} = 1$$

The selection ratios for designs III may be computed by

$$w_{ij}^{\text{III}} = \frac{u_{ij}}{u_{+j}\, p_{i/j}}$$, for habitat type i and animal j (Manly et al. 2002).

Let **T** (I rows × J columns):

$$\mathbf{T} = \left[\sqrt{p_{i/j}}\,(w_{ij}^{\text{III}} - 1)\right]_{i=1\ldots I,\, j=1\ldots J}$$

Then, the factorial analysis of selection ratios for design III data is the eigenanalysis of the triplet (**T**, $\mathbf{I}_{JJ}$, **D**). The origin of the column space, i.e. a row vector of 0 of length I, corresponds to a hypothetical habitat type that is used randomly by all animals. The origin of the row space, i.e. a column vector of 0 of length I, corresponds to a hypothetical habitat type that is used randomly by all animals. The case where a habitat type is not available to an animal, i.e. where $p_{i/j} = 0$, is considered at the end of the appendix. The total inertia of this analysis is equal to

$$\text{Inertia} = \mathbf{T}^{\mathrm{t}}\mathbf{D}\mathbf{T}$$

$$\text{Inertia} = \sum_{i=1}^{I}\sum_{j=1}^{J} p_{i/j} u_{+j} \left(w_{ij}^{\mathrm{III}} - 1\right)^2$$

$$\text{Inertia} = \sum_{i=1}^{I}\sum_{j=1}^{J} p_{i/j} u_{+j} \left(\frac{u_{ij}}{u_{+j} p_{i/j}} - 1\right)^2$$

$$\text{Inertia} = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{\left(u_{ij} - p_{i/j} u_{+j}\right)^2}{p_{i/j} u_{+j}} = \chi^2$$

Thus the inertia of this eigenanalysis is again equal to the standard Chi-square statistic recommended by White and Garrott (1990) to test habitat selection with design III data. This analysis is therefore an extension of the method proposed to analyze design II data.

Note that in some occasions, the availability $p_{i/j}$ may be equal to zero for some animals (e.g. when a given habitat type is absent for some animals whereas it is available for others). In this case, the selection ratio cannot be computed. When this situation occurs, two solutions can be considered.

One first option is to replace the missing values by the mean of the selection ratios for the considered habitat type. However, if the habitat selection varies from one animal to another, then the results may be misleading. For example, the habitat use by an animal may be the result of the habitat types available to it (functional responses, see Mysterud and Ims 1998). If the habitat has been present, the animal may have selected habitat differently. This solution is not recommended, when the aim of the analysis is to establish a typology of animals according to habitat selection.

One second, preferred, option is to replace the missing selection ratios by its expectation under random habitat use, i.e. by setting it equal to 1. Indeed, this analysis maximises the

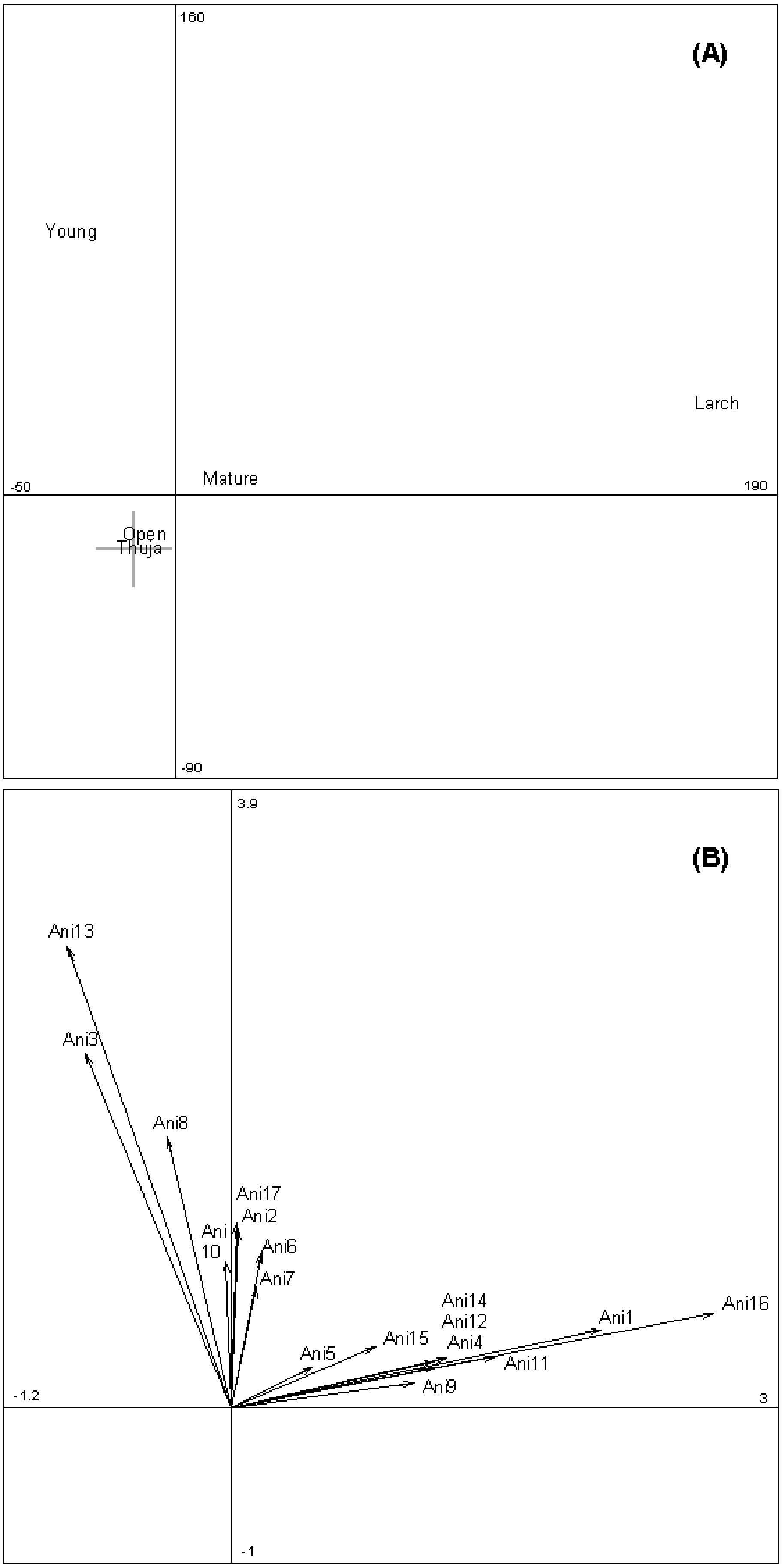
White and Garrott's measure of habitat selection. For a given animal, a habitat type with a selection ratio equal to 1 does not contributes at all to this measure. Therefore, by setting the missing selection ratios equal to 1, one ensures that the missing habitat types do not have any influence in the analysis. As noted in the paper, only the habitat types with large selection ratios for a given animal contribute to the first axes of the analysis.

Table 1. Manly's selection ratio for the 17 squirrels (each with 30 relocations). These measures refer to the selection of the relocations within the study area. The availability of the five habitat types is displayed at the bottom of the table.

	Young	Thuja	Larch	Mature	Open
Ani1	0.00	0.00	16.39	2.21	0.00
Ani2	4.08	0.00	0.00	2.10	0.06
Ani3	9.52	0.00	0.00	0.21	0.43
Ani4	0.45	0.00	9.11	1.79	0.43
Ani5	0.45	0.70	0.00	2.84	0.06
Ani6	3.63	0.00	1.82	1.79	0.25
Ani7	2.72	0.00	0.00	2.21	0.18
Ani8	6.80	0.00	0.00	1.26	0.18
Ani9	0.00	0.00	7.29	2.00	0.43
Ani10	3.63	0.00	0.00	1.79	0.31
Ani11	0.00	0.00	10.93	2.21	0.18
Ani12	0.00	0.00	7.29	2.63	0.06
Ani13	11.79	0.00	0.00	0.32	0.06
Ani14	0.00	0.00	7.29	2.42	0.18
Ani15	0.91	0.00	5.46	2.10	0.31
Ani16	0.00	0.00	23.68	1.68	0.06
Ani17	4.54	0.00	1.82	1.47	0.31
Availability (%)	7.35	4.77	1.83	31.70	54.35

FIGURE LEGENDS

Fig. 1. Results of the factorial analysis of selection ratios carried out to highlight habitat selection by 17 squirrels on five habitat variables. (A) habitat type loadings on the first two factorial axes. The grey cross shows the position of a hypothetical habitat type unused by all animals (B) Animal scores on the first factorial plane.


(A)
160
-50
190
-90
Young
Larch
Mature
Open
Thuja
(B)
3.9
-1.2
3
-1
Ani13
Ani3
Ani8
Ani17
Ani2
Ani
10
Ani6
Ani7
Ani14
Ani12
Ani4
Ani1
Ani16
Ani15
Ani5
Ani11
Ani9

Annexe 3

Article 3 :

“The geographical zonation in the Neotropics of tree species characteristic of the Paraguay-Paraná Basin”

Rodolphe Spichiger, Clément Calenge & Bastian Bise.

Journal of Biogeography (2004), 31, 1489-1501

Journal of Biogeography (*J. Biogeogr.*) (2004) **31**, 1489–1501

ORIGINAL ARTICLE

Geographical zonation in the Neotropics of tree species characteristic of the Paraguay-Paraná Basin

Rodolphe Spichiger, Clément Calenge* and Bastian Bise

Conservatoire et Jardin Botaniques de la Ville de Genève, Ch. de l'Impératrice, Genève, Switzerland

ABSTRACT

Aim This paper uses data from the literature (monographs and taxonomic reviews) to investigate the geographical zonation in South America of 32 common tree species encountered in Paraguay. The actual spatial distribution of the species is then used to provide clues on the plant communities present in the past, and especially during the Wisconsinan age (Last major ice age: 80,000–10,000 BP). The floristic relatedness between the Paraguayan flora and the neighbouring floras is also investigated.

Methods The main vegetation patterns were highlighted using a Discriminant Analysis of the Eigenvectors of Neighbourhood Operator to emphasize the geographical zonation in South America of 32 tree species predominant in the Paraguay-Paraná Basin.

Results Three main vegetation patterns were emphasized: the Chaco, the Paraná-São Francisco gradient and the Peri-Amazonian gradient. The Chaco is well defined. The Peri-Amazonian gradient is characterized by a continuous change in species from the Colombian pole towards the São Francisco pole. The São Francisco gradient shows a continuous change from the São Francisco pole towards the Paraná pole. Some of the species are monocentric (related to only one dispersal centre), whereas others are polycentric (distributed in several poles).

Main conclusions The Chaquean xeromorphic forests are considered as a climax, with a stable composition resulting of the saline soil conditions. The Chaco acts as an edaphic barrier to many species, although the migration of some non-Chaquean species is made possible by the net of gallery-forests and low montane forests between the Río Paraguay and the Andean Piedmont. The Cerrado and the Chaco can be considered as a barrier limiting the expansion of the Paranean forest. This latter formation is a semi-deciduous forest made of a mix of monocentric and polycentric elements. The numerous species found both in Colombia and in Paraná indicate that bridges exist, or have existed, between these two poles. The persistence in the Paraguay-Paraná Basin of a moist forest made of the Paranean assemblages highlighted in this study confirms the hypothesis of moister and cooler phases than today.

*Correspondence: Clément Calenge, UMR CNRS 5558, Laboratoire de Biométrie et Biologie Evolutive, Université Claude Bernard Lyon 1, 43, Boulevard du 11 novembre 1918, 69622 Villeurbanne Cedex, France.
E-mail: calenge@biomserv.univ-lyon1.fr

Keywords

Neotropics, Paraguay, Paraná, Chaco, DAENO, floristic gradients, monocentric species, polycentric species, Pleistocene, Wisconsinan age, Last Glacial Maximum.

INTRODUCTION

The identification of the variation in vegetation composition in response to climate change is a central issue in biogeography, as it provides precious information on the functioning of ecosystems. It helps us to understand the large-scale history of colonization of the plant communities encountered today. Numerous studies have been carried out to determine both the type of climate and the types of flora that occurred during the last ice age in South America, to highlight the processes that have produced the large formations encountered there today (Amazonian forest, Patagonian grasslands, etc.). The Wisconsinan age (80,000–10,000 yr BP), a part of the Pleistocene period, corresponds to the last major ice age, with a global cooling of the climate (see Behling, 1995). However, uncertainty remains concerning the wetness of the climate in South America during this period, and concerning the effect of such a climate on plant communities.

Two main hypotheses are commonly advanced, based on the present distribution of plant species and on palynology studies. According to the first hypothesis, the Wisconsinan stage was characterized by a regression of the forests and an expansion of the tropical and subtropical open formations in South America. For some authors, this Neotropical forest regression was the result of the driest Wisconsinan periods (van der Hammen, 1974, 1982; Ab'Saber, 1982; Ledru, 1993; Servant *et al.*, 1993; van der Hammen & Absy, 1994). For others, and especially palynologists, grassland expansion was a response to both cold and dry phases during the late Quaternary, especially in the south and south-east Brazilian landscape (Markgraf & Bradbury, 1982; Markgraf, 1991; Soubies *et al.*, 1991; Behling, 1993, 1995, 1997a,b, 2002). For Collinvaux, the cooling of the climate was the most important factor leading to forest regression (Collinvaux *et al.*, 1996a,b; Collinvaux, 1997). Finally, according to Mueller (1973) Pleistocenic Neotropical pluvial phases supposedly synchronous with the northern glacial phases, explain the shift of cool-wet adapted taxa (see Spichiger *et al.*, 1995; Pennington, 2000 for reviews).

The alternative hypothesis to the above-mentioned theories is the Residual Pleistocene Seasonally Dry Forest model (RPSDF; Prado & Gibbs, 1993; Pennington, 2000). According to this theory, the drop of temperature and of rainfall levels during the Last Glacial Maximum (LGM) induced the formation of a seasonally dry forest intermingled with rain forest and montane taxa that were confined to the wettest places, in place of the present day (semi-) evergreen Neotropical forests. The remnants of this once much more expanded mesophilous forest, the so-called RPSDF, today form an arc starting from the north-east Brazilian Caatinga ('Caatinga Nucleus', Fig. 1a), reaching the sub-Andean Tucumano-Bolivian forests (Cabrera & Willink, 1973) ('Piedmont Nucleus') and passing through the Paraguay-Paraná-Uruguay Basins ('Misiones Nucleus').

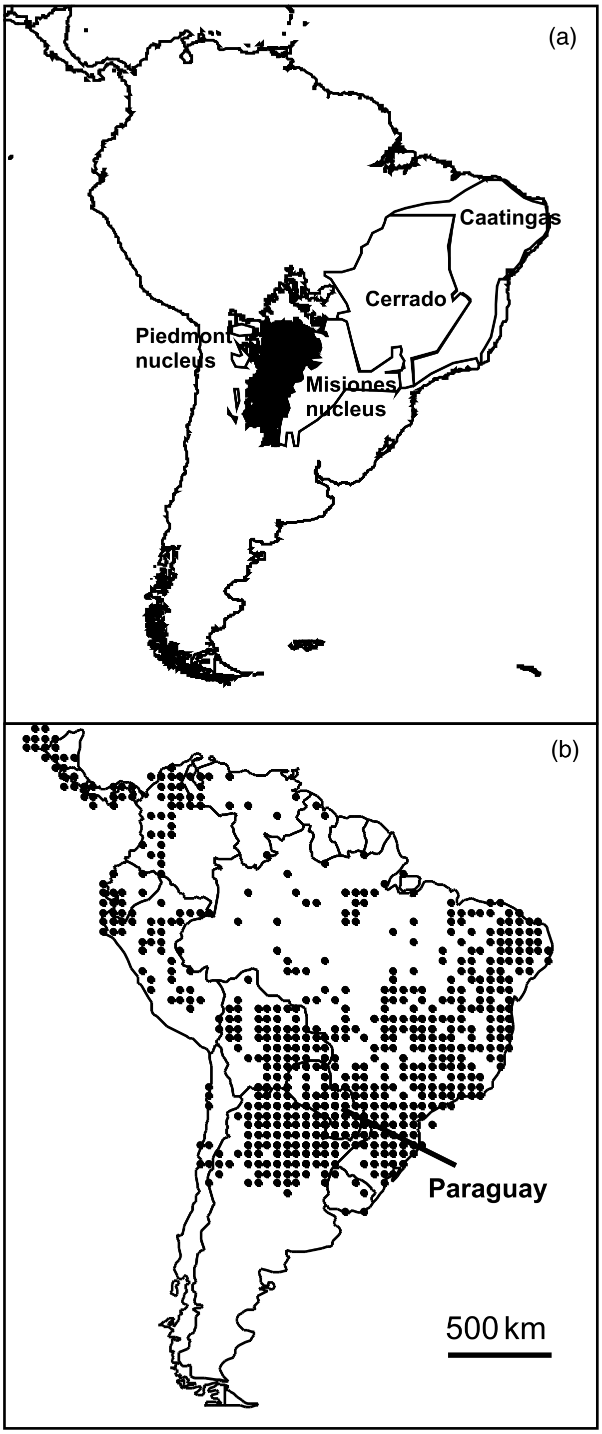


Figure 1 (a) Main biogeographic areas cited in the text, in South America. The dark area corresponds to the Chaco. (b) Distribution of the 1505 occurrences of the 32 tree species under study in South America. The precision of the occurrences location was of 1 degree of both longitude and latitude.

From an ecological and biogeographical point of view, today, the Misiones Nucleus occupies a central place in South America. Indeed, the Paraguayan territory is a huge transition area with a succession of various vegetation types (Spichiger *et al.*, 1995). The great rivers such as Río Paraná, Río Pilcomayo, Río Paraguay and some of their tributaries serve as migration routes for modern floras and faunas. In correlation with climatic and edaphic xericity coming up from the Río Paraná to the Andean Piedmont, the vegetation evolves from semi-deciduous forests in the south-east, to xeromorphic forests and thickets in the north-west. Between these two forest masses extends a wide mosaic of forest-patches, intermingled with Palm-savannas, campos cerrados, fields and cultivated lands. The vegetation mosaic in the Paraguay-Paraná Basin therefore constitutes a huge ecotone where floristic elements of distinct origins are converging or diverging, according to the climate fluctuation (R. Spichiger, C. Calenge & B. Bise, unpubl. data).

Many characteristic and predominant tree species growing in the forests of the Paraguay-Paraná Basin are widespread in the Neotropics (Spichiger *et al.*, 1995), and their present distribution on the continent may provide clues on the plant communities that were encountered in the past. The distribution of some of these species has been used by several authors to produce the RPSDF model (Prado & Gibbs, 1993; Pennington, 2000).

The aims of this research were: (i) to identify the floristic patterns (association between species – communities and gradients) and emphasize the geographical zonation in South America of 32 tree species characteristic of the Paraguayan flora; (ii) based on the above findings, describe the floristic affinities between the Paraguay-Paraná Basin and other areas of South America, with a special focus on ecological and paleoecological explanations.

MATERIAL AND METHODS

Data collection

We studied the spatial distribution in South America of 32 common tree-species encountered in Paraguay (species are listed in the Appendix 1). We chose for this study tree-species that are considered by botanists as predominant and characteristic of the major Paraguayan habitats. For example, we studied *Balfourodendron riedelianum*, because it is a common species along the banks of the Paraná and Uruguay river systems, and is also exploited as a commercial timber species.

We gathered data on spatial distribution of these species from the botanical literature. We especially focused on monographs and taxonomic reviews. These papers are based on the study of herbarium specimens and generally provide distribution maps based on these specimens. The attested presence of a species at some specified location is termed 'occurrence' in the rest of this paper. We digitized these species occurrences on a map of South America, using Arcview GIS (ESRI, 1996). The precision of the locations was of 1 degree of both longitude and latitude (Fig. 1b).

Statistical analysis

Herbarium specimens are often collected in the field without any strict sampling protocol. Consequently, the sampling intensity varies spatially. In other words, inaccessible places are often less sampled than more easily reached places, and the probability of a tree being sampled depends on the location and frequency of this tree. We worked on very common tree species to be confident that the occurrence distribution was as close to the actual distribution of the species as possible. Despite this precaution, varying sampling intensity may imply that the observed distribution of the occurrences (Fig. 1b) does not exactly fit the actual distribution of the trees. However, this is not a limiting factor in our study. Indeed, we aim to emphasize the geographical zonation of the 32 Paraguayan tree species of interest in South America, which is a problem of discrimination of patterns. It is the relative distribution of a species that is of interest, i.e. its distribution relative to the other species. Therefore, we assume that the sampling intensity at a given place is the same for all studied species, although it may vary from one place to another.

We used a discriminant analysis to highlight the geographical zonation of the 32 species in South America. The discriminant analysis is optimal when the objective is to separate groups given measurements for individuals on several variables (Green, 1971; Manly, 1994). This method returns several linear combinations of the descriptive variables for which the ratio (between-group variance)/(total variance) is maximum. In other words, this analysis assigns a numerical score to each occurrence so that the percentage of variation of the scores explained by the factor species is as high as possible.

In this study, we focused on separate tree species according to their geographical range, i.e. discriminate tree species according to some variables that measure the spatial distribution of their occurrences. The eigenvectors of neighbourhood operator provide an interesting way of achieving this objective (Thioulouse *et al.*, 1995). Let $\mathbf{V}$ be a matrix with n rows and n columns (where n is the total number of occurrences). At the intersection of the row i and of the column j, the matrix contains 1 if the ith occurrence is a neighbour of the jth occurrence, and 0 otherwise (a given occurrence is not neighbour of itself; the diagonal of $\mathbf{V}$ thus contains only 0s). The matrix $\mathbf{V}$ is called neighbourhood operator. The eigenvectors of $\mathbf{S}$ are then computed (Méot *et al.*, 1993):

$$\mathbf{S} = \frac{1}{m}\mathbf{D}_n - \frac{1}{m}\mathbf{V},$$

where m is equal to the number of pairs of neighbours (the sum of all values in $\mathbf{V}$), and $\mathbf{D}_n$ a diagonal matrix:

$$\mathbf{D}_n = \mathrm{Diag}(\mathbf{V}\mathbf{1}_n)$$

with $\mathbf{1}_n$ the n-vector of 1. Therefore, at the intersection of the row i and of the column i, $\mathbf{D}_n$ contains the number of neighbours of the occurrence i.

The eigenvectors of $\mathbf{S}$ assign scores to each occurrence so that the score autocorrelation is as high as possible for the study area. These scores may be used to describe the position of the occurrences in the geographical space, as two neighbouring occurrences will have similar scores whereas two distant occurrences will have very different scores.

A Discriminant Analysis of the Eigenvectors of the Neighbourhood Operator (DAENO) by the factor species is best suited to identify the main vegetation patterns in South America. This approach has been recommended by several authors (Méot *et al.*, 1993; C. Calenge, R. Spichiger & D. Chessal, unpubl. data), because of its numerous positive qualities. This method is a discriminant analysis of variables describing the spatial position of the occurrences according to the factor species. These spatial variables (the eigenvectors of the neighbourhood operator) are uncorrelated, which is a highly positive property in discriminant analysis (Manly, 1994). Further, Morellet (1998) has shown that these spatial variables take into account a larger part of the spatial variation in the data than any other type of variables (e.g. polynomial function of Cartesian coordinates). Finally, geographical barriers present in the study area, such as river or roads, may be taken into account by removing the neighbouring relationships between the occurrences located on either side of this barrier (although it was not the case in our study).

We analysed the geographical zonation of the tree species in South America with a DAENO of the tree occurrences within South America. The neighbouring relationships among occurrences were generated using a Delaunay Triangulation algorithm (Renka, 1996), as this is the most common way to generate such relationships (Upton & Fingleton, 1985). We finally tested the pertinence of this analysis using a randomization test (Manly, 1991). At each step of the randomization procedure, a random permutation of the species names was constructed while the rows of the eigenvectors table remained fixed, to simulate a random spatial distribution of the species in South America. The eigenvalues of this analysis measure the percentage of geographical variation explained by the factor species for each factorial axis. Then, the comparison of the eigenvalues of the randomized data sets with the eigenvalues of the observed data set gives the probability that the data set was drawn from random spatial distribution. The map of the occurrences scores may bring useful information on the geographical location of the emphasized vegetation structures. We used a lowess regression (locally weighted regression; Cleveland & Devlin, 1988) to smooth these scores over the study area.

All analyses were carried out using the R software (Ihaka & Gentleman, 1996). We used the R packages 'ade4' and 'adehabitat' for analyses, and the R package 'tripack' for the computation of neighbouring relationships. The R software is freely available on the internet at the URL: http://www.cran.r-project.org/.

RESULTS

Spatial structures

We carried out the DAENO on the 12 first eigenvectors of the neighbourhood operator. The analysis of the data set produces three highly significant factors (61.5% of spatial variation is explained by the factor species on the first axis vs. 38.4% for the second and 32.3% for the third axis; $P < 0.001$). The percentage of explained variance was much lower on the fourth axis (20%), so we restricted our interpretation to the three first axes. The first factorial axis explains a larger part of geographical variation than the following axes. This first factor reveals marked changes in species composition from the Chaco towards other areas in South America (Fig. 2a). The percentage of spatial variation explained by the factor species is similar for the second and third axes (about 35%), which indicated that these two axes should not be interpreted independently. The structures emphasized by these two factors are indeed in two-dimension (Fig. 3a). The second factor highlights a change in species composition along the Peri-Amazonian gradient, from Colombia towards the Río São Francisco area (in the north-east of Brazil, Fig. 2b). The third factorial axis traduces an opposition between the Paraná region and the São Francisco area (Figs 2c & 3a). The floristic variation was continuous along these two axes. In the rest of this paper, we will refer to the structures emphasized by the second and third factor 'Peri-Amazonian gradient' and 'Paraná-São Francisco gradient' respectively.

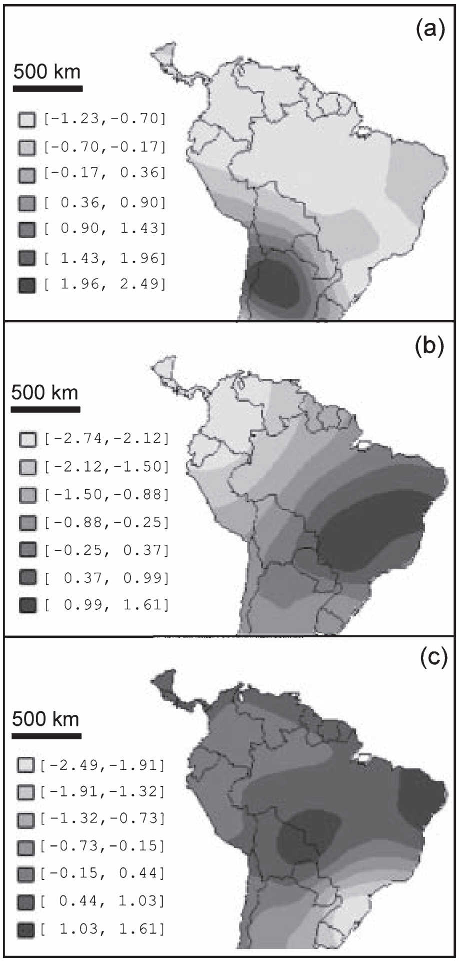


Figure 2 Multispecies spatial patterns obtained by the discriminant analysis of the eigenvectors of neighbourhood operator: (a) first factor, (b) second factor, (c) third factor. For each factor, the occurrence scores are smoothed using lowess regression on 375 neighbours. The grey levels reflect variation in occurrence scores (cf. insert).

Species composition

The Chaquean flora is well defined, as indicated by the species scores on the first axis (Fig. 4). Some species have a large positive score on this first axis, indicating that they are

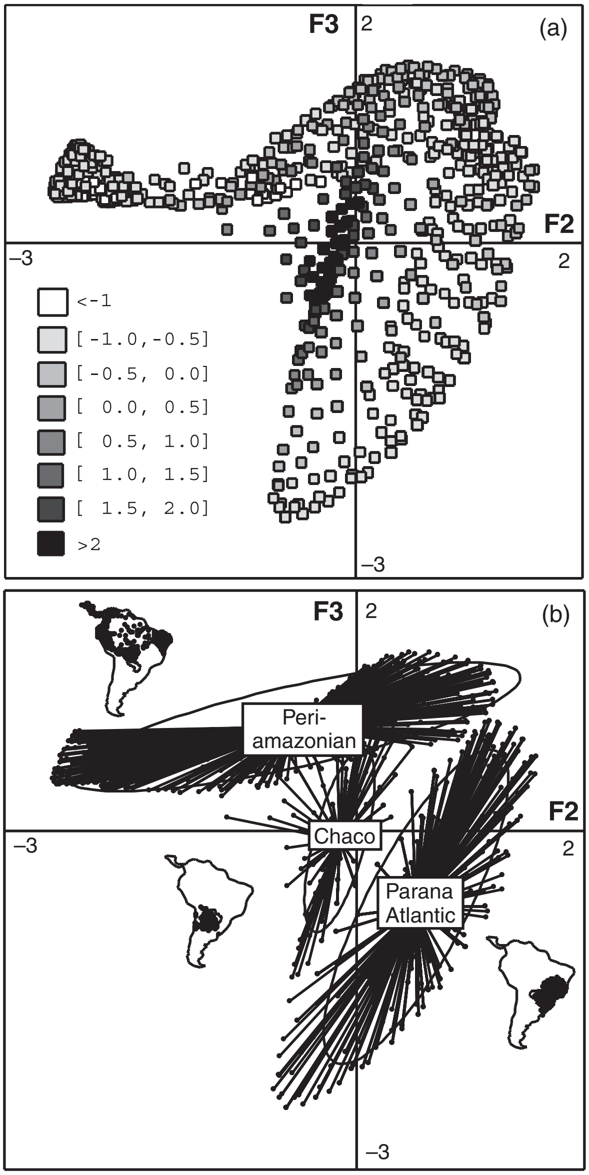


Figure 3 Factorial map of the occurrence scores on the first two axes of the discriminant analysis of the eigenvectors of the neighbourhood operator (abscissa: second axis; ordinate: third axis). (a) The variation in score on the first axis is reflected by grey levels (cf. insert). (b) Categorization of the tree occurrences in three classes according to their position in the three-dimensional space defined by the three first axes. The Chaquean occurrences are identified by a score on the first axis higher than 0.5. The distinction between the Peri-Amazonian gradient and the Paraná Atlantic gradient was made visually (see text). The distribution of tree occurrences in South America is displayed for each structure.

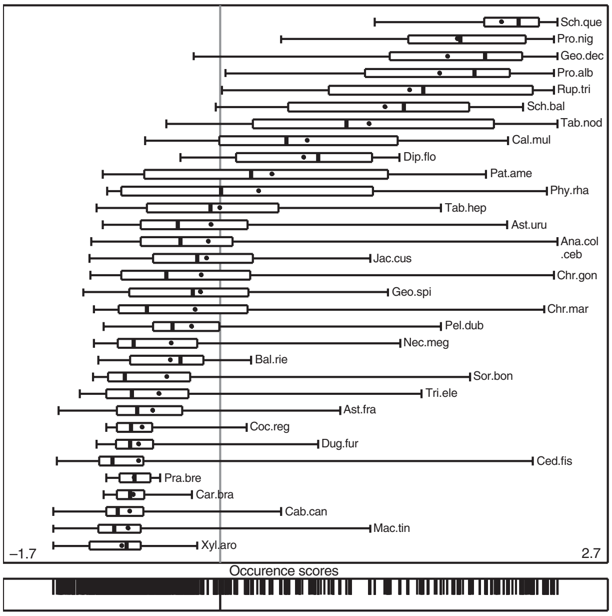


Figure 4 Boxplots displaying the distribution of the tree species on the first factorial axis of the discriminant analysis on the eigenvectors of neighbourhood operator. The upper graph shows the species distribution, and the lower plot presents the scores of the 1505 trees occurrences under study. See Appendix 1 for the species names.

encountered only in the Chaco. Thus, *Schinopsis quebracho-colorado*, *Prosopis nigra*, *Geoffroea decorticans*, *Prosopis alba*, *Ruprechtia triflora* and *Schinopsis balansae* are the most characteristic species. *Tabebuia nodosa*, *Diplokeleba floribunda* and *Calycophyllum multiflorum* are also common in this area, although to a lesser extent than the first group of species. On the other hand, some species with a large negative score on the first axis have rarely or never been reported in the Chaco. This is the case for *Xylopia aromatica*, *Caryocar brasiliense*, *Pradiosa brevipes*, *Balfourodendron riedelianum* or *Cochlospermum regium*.

The variation of the species composition according to the second and third axis is more difficult to interpret, because of the structure highlighted in the 2–3 factorial plane (Fig. 3a). Indeed, the discriminant analysis reveals a complex three-dimensional structure of the scatterpoint of occurrences in a factorial space. We thus artificially categorized this scatterpoint in three classes, to make easier the interpretation of the structures. We defined as Chaquean all occurrences with a score higher than 0.5 on the first axis. We then visually grouped the remaining occurrences into two categories, namely the Peri-Amazonian gradient and the Paraná–São Francisco gradient, according to their position on the factorial plane 2–3 (Fig. 3b).

The Peri-Amazonian gradient is characterized by a continuous change in species from the north of the continent towards the São Francisco area. *Geoffroea spinosa* is uniformly encountered on the whole gradient. In northern South America, this vegetation is mainly made up of *Phyllostylon rhamnoides*, *Trichilia elegans*, *Cedrela fissilis* and *Maclura tinctoria*. These species are also very common in Paraná (see below), which may explain the curved shape of the scatter plot on the factorial plane 2–3 (Fig. 3a). When one moves towards the north-east area of Brazil, tree species more characteristic of the São Francisco area appear progressively. These species include *Xylopia aromatica*, *Astronium fraxinifolium*, *Jacaranda* *cuspidifolia*, *Anadenanthera colubrina* var. *cebil*, *Chrysophyllum gonocarpum*, *Cochlospermum regium*, *Duguetia furfuracea*, *Astronium urundeuva*, *Chrysophyllum marginatum* or *Peltophorum dubium* (Fig. 5a).

The third factorial axis emphasizes a continuous change in vegetation composition from the São Francisco communities on the positive side, towards the Paraná region on the negative side (Fig. 5b). The species characteristic of the São Francisco area (*Astronium fraxinifolium*, *Jacaranda cuspidifolia*, *Cochlospermum regium*, *Duguetia furfuracea*, etc.) become less common with decreasing latitude, and are replaced progressively by the Paranean tree species (*Balfourodendron riedelianum*, *Chrysophyllum gonocarpum*, *Cabralea canjerana*, *Chrysophyllum marginatum*, *Tabebuia heptaphylla*, *Sorocea bonplandii*, *Nectandra megapotamica* and *Patagonula americana*). As stated above, species characteristic of the Colombian assemblages are also common in Paraná (*Trichilia elegans*, *Cedrela fissilis* and *Maclura tinctoria*).

Paraguayan assemblages

Paraguay is thus located at the junction of three major 'affinity poles', namely the São Francisco area, the Paraná region and the Chaco (Fig. 6). The position on the factorial axes of the tree occurrences located within 2 decimal degrees of the Paraguayan boundaries gives an idea of the diversity of floristic communities in Paraguay. Except the Colombian one, all floristic assemblages revealed by the analysis are present in Paraguay (Fig. 7). Thus, Paranean species are encountered in the south-east of the country, whereas the São Francisco assemblage is more common in the north, as revealed by the geographical maps of the occurrences scores on the third axis (Fig. 8b). Moreover, the Chaco flora is found in the western part of Paraguay, as indicated by the maps of occurrence scores on the first axis (Fig. 8a). This confirms that the Paraguay territory may be viewed as a huge ecotone in South America, at the intersection of the Chaco, the São Francisco Basin, and the Paraná Basin.

DISCUSSION

Herbarium data belongs to the large family of 'point pattern data' (e.g. Upton & Fingleton, 1985). The analysis of point pattern data is a major issue in biometry, and the development of new methods therefore has a central place in this field. In this paper, we used the DAENO to highlight vegetation structures in South America. This method is suitable for the analysis of herbarium data, as it answers the stated question, viz. to emphasize the geographical zonation of the tree species according to the spatial distribution of their occurrences. However, further work is required to document more extensively the technical properties of this method as the DAENO has not been widely studied yet. Indeed, several practical questions may arise. For example, we used the very common Delaunay triangulation to generate the neighbouring relationships between the occurrences, but other ways of generating these relationships could also have been employed (e.g. see Gabriel & Sokal, 1969; Pace & Zou, 2000). The sensitivity of the DAENO to this choice is still uncertain. Another question of interest is the sensitivity of this analysis to

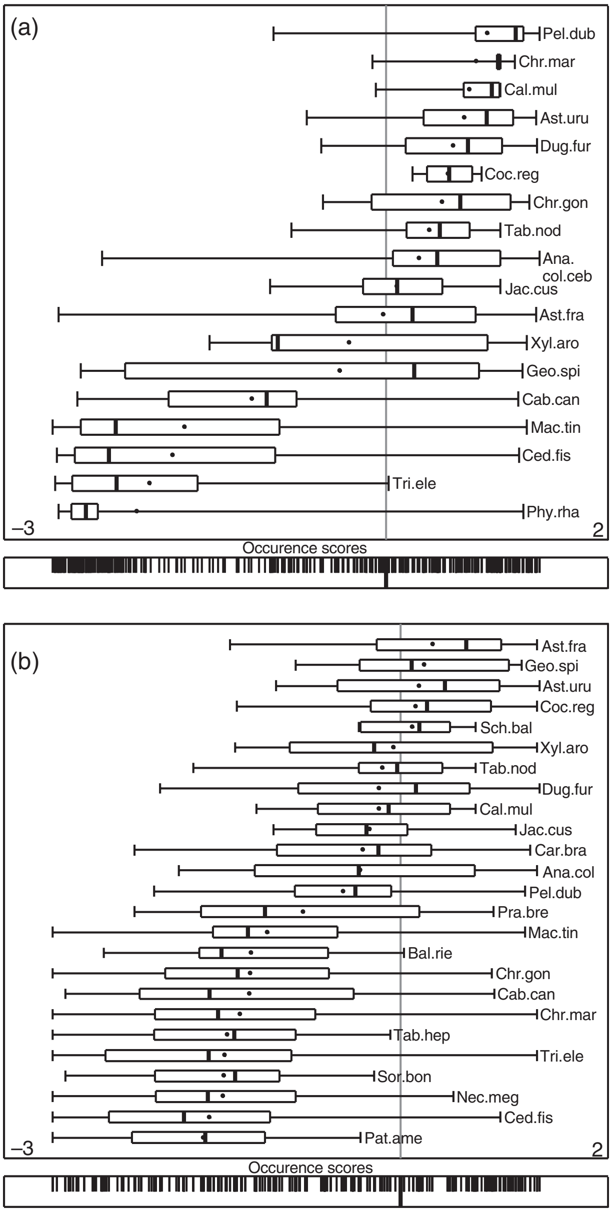


Figure 5 Boxplots displaying the distribution of the tree species on the factorial axes of the discriminant analysis on the eigenvectors of neighbourhood operator. (a) Distribution of the tree occurrences of the Peri-Amazonian gradient on the second factorial axis, (b) distribution of the tree occurrences of the Paraná Atlantic gradient on the third factorial axis. See Fig. 2 for the definition of the gradients. For both plots, the upper graph show the species distribution, and the lower plot presents the scores of the 1505 trees occurrences under study. See Appendix 1 for the species names. Note that some species were absent from either one gradient or the other, and are therefore not represented on the corresponding boxplot, e.g. no occurrence of *Pradiosa brevipes* was noted on the Peri-Amazonian gradient, and this species is therefore not positioned on Fig. 5a.

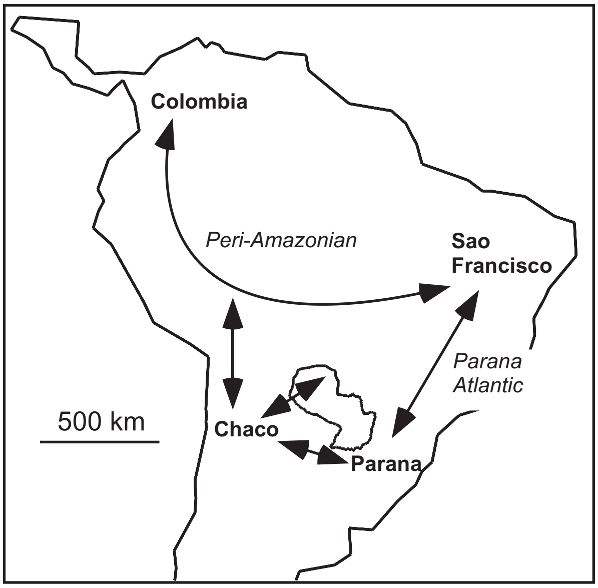


Figure 6 Summary of the structures emphasized by the discriminant analysis of the eigenvectors of neighbourhood operator. The bold names correspond to the four major floristic 'poles' and the main gradients are represented by double arrows labelled in italics. The Paraguay boundary is also displayed.

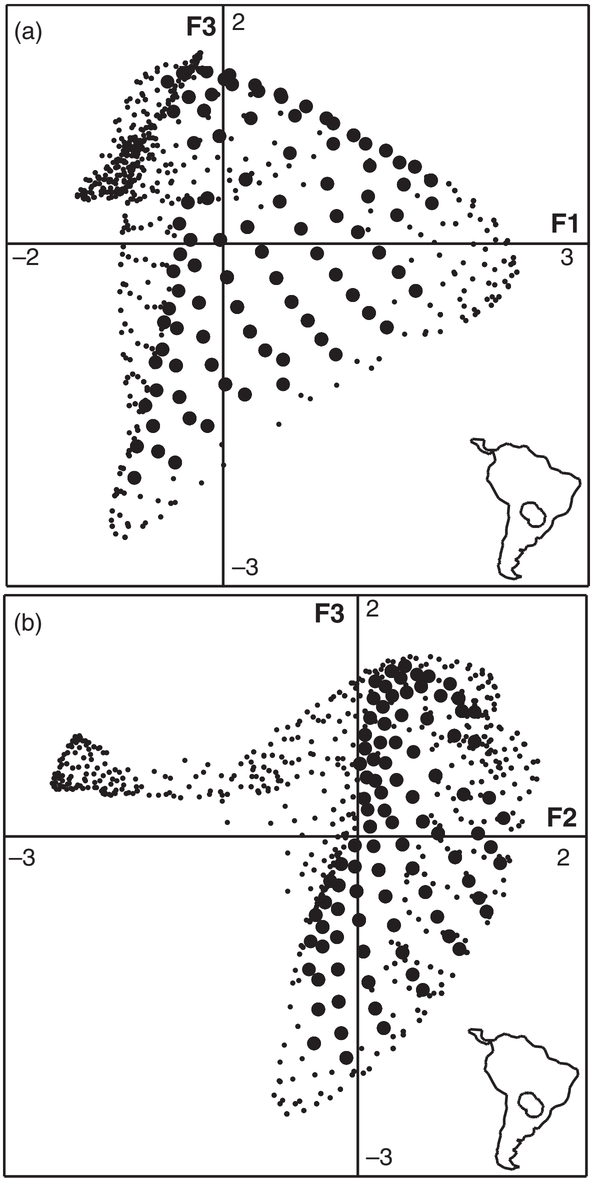


Figure 7 Factorial planes (a) F1-F3 and (b) F2-F3 of the discriminant analysis on eigenvectors of neighbourhood operator. The occurrences located in the area within two decimal degrees of the boundaries of the Paraguay (see maps at the lower right corner of the graphs) are represented by bigger points. Note, that the Paraguayan tree occurrences cover nearly the entire range on the first and the third factors.

the 'form' of the spatial distribution of the occurrences, e.g. when large gaps are found in the occurrences spatial distribution. In our opinion, these questions are minor considerations that are unlikely to affect the validity of the results. The DAENO has proved to be a very robust method of analysis, and gives stable results. The DAENO is a discriminant analysis of variables describing the spatial position of the occurrences by the factor species. Calenge *et al.* (in prep.) have compared three methods relying on this approach that differ only by the way they measure the spatial distribution of the occurrences (Correspondence analysis, canonical correlation trend surface analysis, and the DAENO). They conclude that whatever the method measuring the spatial position of occurrences, the methods relying on this approach give similar results, provided that the underlying structures are sufficiently well defined on the study area. This is the case in our study, as indicated by the strong eigenvalues of the analysis.

Dispersal centres are worked out by plotting the breeding range of species. A dispersal centre is not *a priori* a refuge or a centre of origin. According to Mueller (1973), 'Every species possesses, or used to possess, at least one dispersal centre that was its centre of origin. During the evolution of the taxon, however, the centre of origin and the centre of dispersal can become widely separated from each other.' Using Mueller's concept of 'Dispersal Centre', some of the 32 key species are monocentric (related to only one pole or dispersal centre), whereas most are polycentric (distributed in several poles). These dispersal centres are the Chaco, the Colombia, the Paraná, and the São Francisco region. Three of them reaching the Paraguayan territory (Paraná, Chaco and São Francisco), the position of Paraguay as an ecotone in South America is thus confirmed.

It is difficult to determine, among the 32 focus species, which are monocentric or polycentric. It is quite easy for the species located at one end of the gradients indicated by the analysis (e.g. *Phyllostylon rhamnoides* is related to both the Columbian and Paranean pole, whereas *Patagonula Americana* is only related to the Paranean pole, see Fig. 5). However, determination is more difficult for others that are located at the middle of the gradient (e.g. *Jacaranda cuspidifolia*, mainly

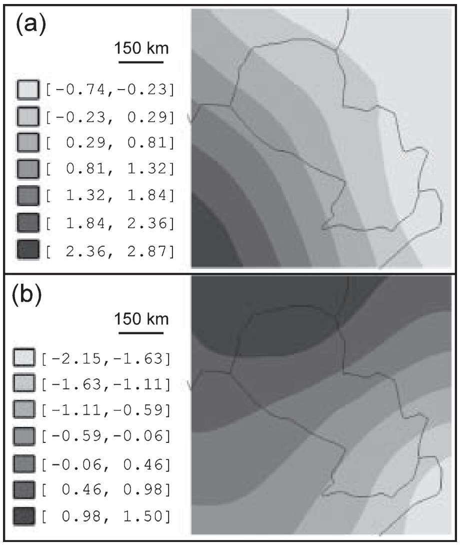


Figure 8 Multispecies spatial patterns obtained by the discriminant analysis of the eigenvectors of the neighbourhood operator within the Paraguay: (a) first factor, (b) third factor. For each factor, the occurrences scores are smoothed using lowess regression on 114 neighbours. The grey levels reflect variation in occurrence scores (cf. insert).

distributed between the Columbian and Sao Francisco Pole). A statistical method is needed to determine whether a species is monocentric or polycentric, based on the distribution of its occurrences.

The Chaquean pole

The Chaco area is the main structure highlighted by the statistical analysis. It is centred in northern Argentina. Previous studies have pointed out that the Chaquean species can be divided into two assemblages (Ramella & Spichiger, 1989; Spichiger *et al.*, 1991, 1995), namely the Chaco Seco (Dry Chaco) and the Chaco Húmedo (Wet Chaco). However, because of the scale of the study and the low precision of the location of tree occurrences in our study (1 degree of longitude × 1 degree of latitude), it is difficult to attain more precise results on Paraguayan communities than those obtained using the maps of factorial axes (Fig. 8), which do not highlight these two communities. This problem of pattern and scale is well known in ecological studies (Levin, 1992). Actually, *Schinopsis quebracho-colorado* and *Ruprechtia triflora* are representatives of the Chaco Seco. *Schinopsis quebracho-colorado* is a most common tree on sandy soils and *Ruprechtia triflora* is a shrub contributing to the understory of the Chaquean xeromorphic forests and thickets (Ramella & Spichiger, 1989; Spichiger *et al.*, 1991). The Chaco Seco pattern is centred in the north-western Chaco, running along the Andean Piedmont and reaching almost the centre of Argentina. It is characterized by the basic loamy-sandy xeric soils, the scarcity of precipitation, and by frost. On the other hand, *Schinopsis balansae*, *Tabebuia nodosa*, *Geoffroea decorticans*, *Prosopis alba* and *P. nigra* are trees characteristic of the Chaco Húmedo (Spichiger *et al.*, 1991; Barberis *et al.*, 2002). These species grow more eastwards than the above-mentioned Chaco Seco species, and are associated with the more humid drainage of the lower Paraguay and to higher levels of precipitation. These elements are less frost-tolerant than those of the Chaco Seco, but are considered generalist because they are found on temporarily waterlogged basic and loamy soils, and on sandier and drier substrates as well (Spichiger *et al.*, in prep.).

The Colombian pole

The Colombian elements have a very characteristic spatial distribution. Most of these species are related to both the Colombian pole and the Paranean pole. This may explain the curved shape of the scatter plot on the factorial plane 2–3 (Fig. 3a). The most characteristic species are *Cedrela fissilis*, *Trichilia elegans*, *Cabralea canjerana* and *Maclura tinctoria*, which are widespread in both the (semi-)evergreen forests and the seasonally dry forests. They grow in all the gallery forests, with the exception of those from the hyper-arid north-eastern Caatinga. They are mostly generalists, reaching their southernmost range in the Paraguay-Paraná Basin and thriving there on well-drained and rich, clayey soils (Oliveira-Filho & Ratter, 1995; Spichiger *et al.*, 1995). *Cedrela fissilis* is widespread in the Neotropics, but is most frequent in the area ranging from southern Brazil through to southern Paraguay and northern Argentina (Pennington, 1981) (see Figs 2c & 5c). According to Klein (1984) this tree grows in all the different types of forests in south-east Brazil, viz. 'Floresta Decidual do Alto Uruguai, Floresta Ombrófila Mista (zona dos pinhais)' and 'Floresta Ombrófila Densa da Costa Atlântica'. It is most commonly found in the 'Floresta Decidual e Semidecidual', i.e. the Paraguayan *Lauraceae* tall semi-deciduous forests (Spichiger *et al.*, 1992). The Colombian pole is also characterized by *Phyllostylon rhamnoides*, a good indicator of local edaphic humidity, which also occurs in the Chaco, in fact, in the wettest places of the Chaco Húmedo (Spichiger *et al.*, 1991). Finally, *Geoffroea spinosa* is related to three dispersal centres: Colombia, São Francisco and Chaco (the Chaco Húmedo). It is characterized by one of the widest ecological amplitudes among the analysed key species (Spichiger et al., in prep.).

The São Francisco pole

The São Francisco pole is built up by '*Residual Pleistocene Seasonally Dry Flora*' elements (*Astronium urundeuva*, *Anadenanthera colubrina* var. *cebil*, *Peltophorum dubium*, etc.) (Prado & Gibbs, 1993; Pennington, 2000) and by Cerrado-related species (*Astronium fraxinifolium*, *Jacaranda cuspidifolia*, *Cochlospermum regium*, *Duguetia furfuracea*). The São Francisco area is a mosaic of savannas and gallery-forests that support a drier climate than the Paraná and the Colombian areas (Intergovernmental Panel on Climate Change, 2003). The galleries of the São Francisco region act as migration routes between the Paraná and Colombian poles (Oliveira-Filho & Ratter, 1995), as does the gallery net of the Chaco Húmedo (Spichiger *et al.*, 1991, 1995). Most of these gallery species are generalists related to the Paraná and the other poles as well. Some species are found along a continued meridional arch from the Andean Piedmont to the São Francisco pole, through the Paraguay-Paraná Basin (e.g. *Anadenanthera colubrina* var. *cebil*).

The Paranean pole

Sorocea bonplandii and *Nectandra megapotamica* are monocentric being related only to the Paraná dispersal centre. Moreover, a lot of polycentric elements are strongly related to the Paraná dispersal centre: *Balfourodendron riedelianum*, *Chrysophyllum gonocarpum*, *Chrysophyllum marginatum*, *Peltophorum dubium*, *Tabebuia heptaphylla*. *Astronium urundeuva*, *Xylopia aromatica*, *Peltophorum dubium*, *Patagonula americana*. The monocentric elements, together with polycentric elements are strongly represented in the area that constitutes the Paranean semi-deciduous forest (termed Floresta Ombrofila Mista in the Brazilian literature). This forest is centred in the Upper Paranean-Upper Uruguayan Basins. It extends from north-east Argentina to south-east Brazil and to the southern Planalto, along the Paraná river and its tributaries. Due to the high frequency of *Lauraceae* species, these forests are called Monte de Laurel, or Laurisilva (Spichiger *et al.*, 1995), or *Parapiptadenia rigida–Nectandra saligna* forests by the Argentineans (Eskuche, 1982, 1984). These forests belong to the Paraneense Province according to Cabrera & Willink (1973), and more or less to the Prado's Misiones Nucleus and to the Paranean elements according to Spichiger *et al.* (1995). The forests found on the well-drained soils, i.e. on the uplands, are rich in Meliaceae and Sapotaceae, while in contrast those growing on ill-drained substrates are richer in Sapindaceae and Tiliaceae (Spichiger *et al.*, 1992). Moreover, the Paraná-related elements are also scattered in the forest-islets and gallery-forests constituting the forest-savanna mosaics on the western bank of the Río Paraguay and at the confluence between the Río Paraguay and Río Paraná. Bernardi (1984) pinpointed the similarities between the dendrofloras extending from eastern Paraguay up to the Río São Francisco, and the separation between the Amazon Basin and the Paraná drainage system.

Paleoecological hypotheses

It is difficult to draw conclusions about past climates from present day distributions of plant species. Even when the fruits are not capable of rapid dispersal the distribution range of a given species may change markedly over a 10,000-year period (as is the case for post-glaciation colonization by trees in Europa). However, present day distributions may contribute elements of information that confirm or deny existing hypotheses based on results obtained in the fields of climatology and palynology (Markgraf & Bradbury, 1982; Markgraf, 1991; Soubies *et al.*, 1991; Behling, 1993, 1995, 1997a,b, 2002; Collinvaux *et al.*, 1996a,b; Collinvaux, 1997), or in studies of the distribution of several species considered separately (Prado & Gibbs, 1993; Spichiger *et al.*, 1995; Pennington, 2000). The novel aspect of our study is that it relies on multivariate analysis of herbarium data. We thus consider the issue from a different perspective using data that can bring new information to the problem.

The Andean uplift during the Oligocene provoked the formation of a syncline, which was first occupied by the sea, leaving, after it withdrew, marine salty sediments on which Andean deposits accumulated (Putzer, 1962). This most particular substratum is now colonized by a xeromorphic flora, the Chaquean flora, extending from north-west Argentina up to north-west Paraguay and south-east Bolivia. Thus the Paraguayan xeromorphic forest with 'quebrachos' (*Aspidosperma quebracho-colorado* and *Schinopsis balansae*) has to be considered as a climax resulting of the saline soil conditions and, perhaps, as extra-tropical formation because of the relatively high number of frost-days in the area (Spichiger *et al.*, 1991, 1995; Prado, 1993a,b). The Chaco biome is thus likely a Tertiary or early Pleistocene relict (this hypothesis is, however, challenged by Mueller (1973) who considers the Chaquean biome to be much younger). The salty, hyper-arid or waterlogged soils of the Chaco appear as an unpassable edaphical barrier for the most of the semi-deciduous species. This barrier can be overstepped by some Paraná-related elements (*Tabebuia heptaphylla*, *Astronium urundeuva*, *Anadenanthera colubrina* var. *cebil*, etc.) because of the net of gallery-forests oriented east–west, or because of the cerros, ills and cordilleras, where the soil conditions are different (Spichiger *et al.*, 1991).

Wisconsinan drier periods have been postulated by many authors (e.g. Absy *et al.*, 1991; van der Hammen & Absy, 1994; Behling, 2002). Thus, Prado & Gibbs (1993) and Pennington (2000) states that during the LGM (25,000–15,000 yr BP) seasonally dry forests (the so-called Residual Pleistocene Seasonally Dry Flora), and not savannas, were found in place of the present (semi-)evergreen Amazonian forests. The present-day patterns of distribution may thus be remnants of a once much more expanded distribution that survive presently only at the edges of the Amazon Basin. For instance, the present day huge circum-Amazonian occurrence of *Cedrela fissilis* could be explained by a Holocene (*c.* 10,000 yr BP) rise in both temperature and rainfall. This climate variation may have caused the drought-tolerant and generalist species such as *C. fissilis* to move to the edges of the Amazon Basin and be replaced in the wettest areas by drought-sensitive elements. This hypothesis of generalist elements dominant in the Amazon Basin during the LGM at the expense of the present Amazonian elements is corroborated by the present-day continuous or disjunct circum-Amazonian pattern of distribution of the polycentric elements found in the Colombian, São Francisco and Paraná poles.

The Paraná flora is composed of a mixture of monocentric and polycentric species. This forest is a semi-deciduous forest that is moister (Intergovernmental Panel on Climate Change,

2003) and distinct from the other Neotropical seasonally dry forests. It seems likely that the monocentric Paranean elements or the polycentric Paraná-related species would have diverged from or converged towards the Paraná area according to the climate changes. During the LGM, the drought-sensitive species, such as *Nectandra megapotamica* or *Sorocea bonplandii*, would have probably withdrawn towards moister areas, such as the Atlantic coast. On the other hand, a south-Amazonian arch pattern of distribution (*Anadenanthera colubrina* var. *cebil* among others) show that bridges exist, or have existed, between the Paraná Basin and the Colombian pole (the so-called Tucumano-Bolivian forests of the Piedmont area). Gregarious populations of *A. colubrina* var. *cebil* build up the vegetation of the montane forests of the Cerro León, the Cerro Cabrera (west Paraguay; Spichiger *et al.*, 1991) and the Tucumano-Bolivian forests. *Calycophyllum multiflorum* is a leading element in the Tucumano-Bolivian forests, in the Paraná area and in the gallery-forests of the Chaco Húmedo. We can infer that extra-zonal semi-deciduous forests in the Chaco connect Andean Piedmont with the Paraná Basin. Oliveira-Filho & Ratter (1995) supported the hypothesis of the expansion of Paranense semi-deciduous forests via their penetration into the Cerrado Province via the basalt-derived fertile soils of western Minas Gerais and southern Goías (central Brazil). Some of them could have extended westwards along the Andean Piedmont, establishing links with the Colombian pole.

Brown (1982), on the basis of palaeoecological data on butterflies, indicates the Iguaçu region (northern Paraná) as a probable forest refuge for butterflies during the last 20,000 years. Ray & Adams (2001) infer from GIS models the presence of a temperate evergreen forest in Upper Paraná during the LGM surrounded by xeric grasslands. The subsistence in the Paraguay-Paraná Basin of a moist forest made of the Paranean assemblages highlighted in this study supports these propositions. Thus, the hypothesis of a Quaternary forest refuge in the Upper Paraná area during the late Pleistocene or the early Holocene is not to be rejected. The hypothesis is corroborated by the present-day Paraná assemblage and its relative isolation from other Neotropical forests by both the Chaco Seco and the Cerrado biomes. We can also infer from frost-day occurrences in the Paraná valley, as well as from *Araucaria* patches in the Paraguayan Alto Paraná (northern Paraná; Spichiger *et al.*, 1992), that an adaptation to a cooler and moister climate as today has taken place. According to Klein (1975), 20,000 years ago the Paranean forest was a mixture of lowland semi-deciduous taxa and moist-temperate elements (*Araucaria* and related species). Therefore, the interfluvial area between the Ríos Paraná and Ríos Paraguay would have been occupied by species assemblages similar to those that are encountered today in the meridional Brazilian Planalto. For several authors (e.g. Iriondo & Garcia, 1993), Paraná elements, more or less intermingled with temperate forest species and grassland taxa (*Araucaria*, *Podocarpus*, *Drimys* and Poaceae), should have been much more widespread southwestwards during the LGM. During a cooler and moister climate, the elements of the so-called Paraneense Province – probably mixed with cool-adapted rain forest elements and grassland – should have been much more widespread than they are today (Cabrera & Willink, 1973; Spichiger *et al.*, 1992).

In this paper, we focused only on the spatial distribution of 32 tree species. However, it may also be of interest to emphasize more precisely the relationships between the tree species and their environment. Guisan & Zimmermann (2000) have described several modelling techniques that can be used to draw conclusions on such relationships. We also need to investigate whether herbarium data can be used to fit such a model. Indeed, this approach implies that the observed distribution of the habitat characteristics for a given species is representative of the actual distribution. This hypothesis may be reasonable even when the observed geographical distribution of the occurrences is not representative of the actual geographical distribution of a given species. The ecological representativeness of the herbarium data first needs to be evaluated. In the case where this hypothesis is considered reasonable, a model predicting the potential distribution of each species according to environmental and climatic variables would allow us to test the effect of various climatic scenarios on the distribution of each species of interest. This would therefore bring additional arguments that confirm or deny existing paleoecological hypotheses concerning the climate during the LGM.

ACKNOWLEDGMENTS

We would like to warmly thank Daniel Chessel (University of Lyon) for his most valuable help in statistical analyses, Anne-Beatrice Dufour (University of Lyon) for correcting an earlier version of this paper, as well as Michèle Price (Conservatoire et Jardin Botaniques de Genève) for correcting the English. We are also grateful to two anonymous referees for their constructive and helpful comments, which greatly improved this article.

REFERENCES

Ab'Saber, A.N. (1982) The paleoclimate and paleoecology of Brazilian Amazonia. *Biological diversification in the tropics* (ed. by G.T. Prance), pp. 41–59. Columbia University Press, New York.

Absy, M.L., Cleef, A., Fournier, M., Martin, L., Servant, M., Sifeddine, A., Ferreira da Silva, M., Soubies, F., Suguio, K., Turcq, B. & van der Hammen, T. (1991) Mise en évidence de quatre phases d'ouverture de la forêt dense dans le sud-est de l'Amazonie au cours des 60000 dernières années. Première comparaison avec d'autres régions tropicales. *Comptes Rendus de l'Académie des Sciences, Paris*, **312,** 673–678.

Barberis, I.M., Batista, B.W., Pire, E.F., Lewis, J.P. & Leon, R.J.C. (2002) Woody population distribution and environmental heterogeneity in a Chaco forest, Argentina. *Journal of Vegetation Science*, **13,** 607–614.

Behling, H. (1993) Untersuchungen zur spätpleistozänen und holozänen Vegetations- und Klimageschichte der tropischen Küstenwälder und der Araukarienwälder in Santa Catarina (Südbrasilien). *Dissertationes Botanicae*, **206,** 1–149.

Behling, H. (1995) A high resolution Holocene pollen record from Lago do Pires, SE Brazil: vegetation, climate and fire history. *Journal of Paleolimnology*, **14,** 253–268.

Behling, H. (1997a) Late quaternary vegetation, climate and fire history from the tropical mountain region of Morro de Itapeva, SE Brazil. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **129,** 407–422.

Behling, H. (1997b) Late Quaternary vegetation, climate and fire history in the Araucaria forest and campos region from Serra Campos Gerais (Paraná), S Brazil. *Review of Palaeobotany and Palynology*, **97,** 109–121.

Behling, H. (2002) Carbon storage increases by major forest ecosystems in tropical south America since the Last Glacial Maximum and the early Holocene. *Global and Planetary Change*, **33,** 107–116.

Bernardi, L. (1984) Contribución a la dendrología Paraguaya. Primera parte. Apocynaceae – Bombacaeceae – Euphorbiaceae – Flacourtiaceae – Mimosoideae – Caesalpinioideae – Papilionatae. *Boissiera*, **35,** 1–341.

Brown, K.S. (1982) Paleoecology and regional patterns of evolution in forest butterflies. *Biological diversification in the tropics* (ed. by G.T. Prance), pp. 255–308. Columbia University Press, New York.

Cabrera, A.L. & Willink, A. (1973) *Biogeogragrafía de América Latina*. Secretarial General OEA, Washington, DC.

Cleveland, W.S. & Devlin, S.J. (1988) Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association*, **83,** 596–610.

Collinvaux, P. (1997) *The Ice-Age Amazon and the problem of diversity*. NOW/Huygenslezing 1997, The Hague, the Netherlands.

Collinvaux, P.A., de Oliveira, P.E., Moreno, J.E., Miller, M.C. & Bush, M.B. (1996a) A long pollen record from lowland Amazonia: forest and cooling in glacial times. *Science*, **274,** 85–88.

Collinvaux, P.A., Liu, K.-B., de Oliveira, P.E., Bush, M.B., Miller, M.C. & Steinitz Kannan, M. (1996b) Temperature depression in the lowland tropics in glacial times. *Climatic Change*, **32,** 19–33.

Eskuche, U. (1982) Struktur und Wirkungsgefüge eines subtropischen Waldes Südamerikas. *Struktur und Dynamik von Wäldern* (ed. by H. Dierschke), pp. 49–64. Berichte der Internationalen Symposien der Internationalen Vereinigung Vegetationskunde. J. Cramer, Vaduz.

Eskuche, U. (1984) Vegetationsgebiete von Nord- und Mittelargentinien. *Phytoceonologia*, **12,** 185–199.

ESRI (1996) *Using Arcview GIS. The geographic information system for everyone*. Environmental Systems Research Institute, Inc, Redlands, CA.

Gabriel, K.R. & Sokal, R.R. (1969) A new statistical approach to geographic variation analysis. *Systematic Zoology*, **18,** 259–278.

Green, R.H. (1971) A multivariate statistical approach to the hutchinsonian niche: bivalve molluscs of central Canada. *Ecology*, **52,** 543–556.

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

van der Hammen, T. (1974) The Pleistocene changes of vegetation and climate in tropical South America. *Journal of Biogeography*, **1,** 3–26.

van der Hammen, T. (1982) Paleoecology of tropical South America. *Biological diversification in the tropics* (ed. by G.T. Prance), pp. 60–66. Columbia University Press, New York.

van der Hammen, T. & Absy, M.L. (1994) Amazonia during the last glacial. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **109,** 247–261.

Ihaka, R. & Gentleman, R. (1996) R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, **5,** 299–314.

Intergovernmental Panel on Climate Change (2003) The IPCC Data Distribution Centre. http://www.ipcc-ddc.cru.uea.ac.uk/.

Iriondo, M.H. & Garcia, N.O. (1993) Climatic variations in the Argentine plains during the last 18000 years. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **101,** 209–220.

Klein, R.M. (1975) Southern Brazilian phytogeographic features and the probable influence of upper Quaternary climatic changes in the floristic distribution. *Boletim Paranáense Geoscienca*, **33,** 67–88.

Klein, R.M. (1984) Meliáceas, *Flora illustrada Catarinense* (ed. by P.R. Reitz), pp. 103–113. Itajaí, Santa Catarina, Brasil.

Ledru, M.P. (1993) Late Quaternary environmental and climatic changes in Central Brazil. *Quaternary Research*, **39,** 90–98.

Levin, S.A. (1992) The problem of pattern and scale in ecology. *Ecology*, **73,** 1943–1967.

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

Manly, B.F.J. (1994) *Multivariate statistical methods. A primer*, 2nd edn. Chapman & Hall, London.

Markgraf, V. (1991) Younger Dryas in southern South America? *Boreas (Oslo)*, **20,** 63–69.

Markgraf, V. & Bradbury, J.P. (1982) Holocene climatic history of South America. *Striae*, **16,** 40–45.

Méot, A., Chessel, D. & Sabatier, R. (1993) Opérateurs de voisinage et analyse des données spatio-temporelles. *Biométrie et environnement* (ed. by J.D. Lebreton & B. Asselain), pp. 45–71. Masson, Paris.

Morellet, N. (1998) *Des outils biométriques appliqués aux suivis des populations animals: l'example des cervidés*. PhD thesis, University of Claude Bernard Lyon 1, Villeurbanne, France.

Mueller, P. (1973) *The dispersal centres of terrestrial vertebrates in the Neotropical realm: a study in the evolution of the Neotropical biota and its native landscapes*. Junk, La Hague.

Oliveira-Filho, A.T. & Ratter, J.A. (1995) A study of the origin of central Brazilian forests by the analysis of plant species distribution patterns. *Edinburgh Journal of Botany*, **52,** 141–194.

Pace, R.K. & Zou, D. (2000) Closed-form maximum likelihood estimates of nearest neighbor spatial dependence. *Geographical Analysis*, **32,** 154–172.

Pennington, T.D. (1981) Meliaceae. *Flora Neotropica*, **28,** 1–474.

Pennington, T.D. (2000) Neotropical seasonally dry forests and Quaternary vegetation changes. *Journal of Biogeography*, **27,** 261–273.

Prado, D.E. (1993a) What is the Gran Chaco vegetation in South America? I. A review. Contribution to the study of the flora and vegetation of the Chaco. *Candollea*, **48,** 145–172.

Prado, D.E. (1993b) What is the Gran Chaco vegetation in South America? II. A redefinition. Contribution to the study of the flora and vegetation of the Chaco. *Candollea*, **48,** 615–629.

Prado, D.E. & Gibbs, P.E. (1993) Patterns of species distributions in the dry seasonal forests of South America. *Annals of the Missouri Botanical Garden*, **80,** 902–927.

Putzer, H. (1962) *Geologie von Paraguay*. Gebüder Bornträger, Berlin-Nikolassee.

Ramella, R. & Spichiger, R. (1989) Interpretación preliminar del medio físico y de la vegetación del Chaco boreal. Contribucióacn al estudio de la flora y de la vegetación del Chaco. *Candollea*, **44,** 639–680.

Ray, N. & Adams, J.M. (2001) A GIS-based vegetation map of the world at the Last Glacial Maximum (25,000–15,000 BP). Internet archaeology (http://www.intarch.ac.uk/journal/issue11/rayadams_toc.html).

Renka, R.J. (1996) Algorithm 751: TRIPACK: a constrained two-dimensional {Delaunay} triangulation package. *ACM Transactions on Mathematical Software*, **22,** 1–8.

Servant, M., Maley, J., Turcq, B., Absy, M.L., Brenac, P., Fournier, M. & Ledru, M.P. (1993) Tropical forest changes during the late Quaternary in African and South American lowland. *Global Planet Change*, **7,** 25–40.

Soubies, F., Suguion, K., Martin, L., Leprun, J., Servant, M., Turcq, B., Fournier, M., Delaune, M. & Sfeddine, A. (1991) The Quaternary deposits of the Serra dos Carajas (State of Pará, Brazil) – ages and other preliminary results. Proceedings Global Changes in South America during the Quaternary. *Boletim Instituto de Geosciencias-Universidade de São Paulo. Publicação Especial*, **8,** 223–243.

Spichiger, R., Ramella, L., Palese, R. & Mereles, F. (1991) Proposición de leyenda para la cartografía de las formaciones vegetales del Chaco paraguayo. Contribución al estudio de la flora y de la vegetación del Chaco. *Candollea*, **46,** 541–564.

Spichiger, R., Bertoni, B.S. & Loizeau, P.-A. (1992) The forests of the Paraguayan Alto Paraná. *Candollea*, **47,** 219–250.

Spichiger, R., Palese, R., Chautems, A. & Ramella, L. (1995) Origin, affinities and diversity hot spots of the Paraguayan dendrofloras. *Candollea*, **50,** 515–537.

Thioulouse, J., Chessel, D. & Champely, S. (1995) Multivariate analysis of spatial patterns: a unified approach to local and global structures. *Environmental and Ecological Statistics*, **2,** 1–14.

Upton, G. & Fingleton, B. (1985) *Spatial data analysis by example. Vol. 1: Point pattern and quantitative data*. John Wiley and Sons, Chichester.

BIOSKETCHES

Rodolphe Spichiger is Professor of Botany at the University of Geneva (Switzerland) and Head of the Botanical Garden of Geneva. He received his PhD from the University of Geneva in 1977. His research interest is the flora and vegetation of the tropics, especially of South America and Africa.

Clément Calenge is a PhD student at the Department of Biometry, University of Lyon (France). He is working on statistical methods intended to analyse point pattern in geographical space.

Bastian Bise is a geography student at the University of Geneva (Switzerland). He works for the Botanical Garden of Geneva on the effect of environmental variables on the spatial distribution of tree species in South America.

Appendix 1 List of names and abbreviations (abbr.) used for the 32 tree species under study in South America. The number of occurrences for each species is also indicated

Species name	Abbr.	Number
Anadenanthera colubrina var. *cebil* (Griseb.) Altschul	Ana.col.ceb	115
Astronium fraxinifolium Schott	Ast.fra	73
Astronium urundeuva (Allemão & M. Allemão) Engl.	Ast.uru	56
Balfourodendron riedelianum (Engl.) Engl.	Bal.rie	15
Cabralea canjerana (Vell.) Mart.	Cab.can	48
Calycophyllum multiflorum Griseb.	Cal.mul	27
Caryocar brasiliense Cambess.	Car.bra	34
Cedrela fissilis Vell.	Ced.fis	88
Chrysophyllum gonocarpum (Mart. & Eichler) Engl.	Chr.gon	54
Chrysophyllum marginatum (Hook. & Arn.) Radlk.	Chr.mar	60
Cochlospermum regium (Schrank) Pilg.	Coc.reg	34
Diplokeleba floribunda N. E. Br.	Dip.flo	11
Duguetia furfuracea (A. St.-Hil.) Benth. & Hook. f.	Dug.fur	88
Geoffroea decorticans (Hook. & Arn.) Burkart	Geo.dec	65
Geoffroea spinosa (Willd.) Morong	Geo.spi	41
Jacaranda cuspidifolia Mart.	Jac.cus	38
Maclura tinctoria (L.) D. Don ex Steud.	Mac.tin	159
Nectandra megapotamica (Spreng.) Mez	Nec.meg	47
Patagonula americana L.	Pat.ame	25
Peltophorum dubium (Spreng.) Taub.	Pel.dub	40
Phyllostylon rhamnoides (J. Poiss.) Taub.	Phy.rha	28
Pradosia brevipes (Pierre) T.D. Penn.	Pra.bre	14
Prosopis alba Griseb.	Pro.alb	48
Prosopis nigra (Griseb.) Hieron.	Pro.nig	33
Ruprechtia triflora Griseb.	Rup.tri	9
Schinopsis balansae Engl.	Sch.bal	24
Schinopsis quebracho-colorado (Schltdl.) F. A. Barkley & T. Mey.	Sch.que	28
Sorocea bonplandii (Baill.) W.C. Burger, Lanj. & Wess. Boer	Sor.bon	39
Tabebuia heptaphylla (Vell.) Toledo	Tab.hep	38
Tabebuia nodosa (Griseb.) Griseb.	Tab.nod	59
Trichilia elegans A. Juss.	Tri.ele	49
Xylopia aromatica (Lam.) Mart.	Xyl.aro	18

Annexe 4

Article 4 :

“The discriminant analysis of the spatial distribution of vegetal species occurrences: I. theoretical aspects”

Clément Calenge, Rodolphe Spichiger, Daniel Chessel & Cyrille Chatelain.

Candollea, sous presse

Candollea 60(2): 000-000 (2005)

Discriminant analysis of the spatial distribution of plant species occurrences: I. Theoretical aspects

CLÉMENT CALENGE,
RODOLPHE SPICHIGER,
DANIEL CHESSEL
&
CYRILLE CHATELAIN

ABSTRACT

CALENGE, C., R. SPICHIGER, D. CHESSEL & C. CHATELAIN (2005). Discriminant analysis of the spatial distribution of plant species occurrences: I. Theoretical aspects. *Candollea* 60: 000-000. In English, English and French abstracts.

This paper is the first of two articles describing theoretical and practical aspects of spatial discrimination of species distributions in a given area. Three multivariate methods used to show species geographical zonation from lists of species occurrence data from herbarium records are discussed here. These methods are (i) *Correspondence Analysis* (CA), (ii) *Canonical Correlation Trend Surface Analysis* (CCTSA), and (iii) *Discriminant Analysis on Eigenvectors of Neighbourhood Operator* (DAENO). Their use is illustrated through the analysis of the spatial distribution of three virtual species. This paper shows that these methods are forms of *Discriminant Analysis* (DA) which use the spatial position of the species occurrences as variables; they just differ in the way they measure this position in space. It is concluded that these three methods are likely to produce the same results from a given data set, providing that the underlying spatial structures are well defined within the study area.

RÉSUMÉ

CALENGE, C., R. SPICHIGER, D. CHESSEL & C. CHATELAIN (2005). L'analyse discriminante de la distribution spatiale d'occurrences d'espèces végétales: I. Aspects théoriques. *Candollea* 60: 000-000. En anglais, résumés anglais et français.

Cet article est le premier d'une série de deux articles décrivant les aspects théoriques et pratiques de la discrimination spatiale de distributions d'espèces dans une zone donnée. Trois méthodes multivariées sont discutées ici et qui peuvent être utilisées pour mettre en évidence la zonation géographique d'espèces à partir de listes d'occurrences d'espèces issues de données d'herbier. Ces méthodes sont: (i) l'*Analyse Factorielle des Correspondances* (AFC), (ii) l'*Analyse Canonique des Corrélations Appliquées aux Tendances de Surface* (ACCATS), et (iii) l'*Analyse Discriminante sur Vecteurs Propres du Graphe de Voisinage* (ADVPGV). Leur utilisation est illustrée par la distribution spatiale de trois espèces virtuelles. Cet article montre que ces méthodes sont des formes d'*Analyse Discriminante* (AD) qui utilisent comme variables la position spatiale des occurrences d'espèces, et qu'elles ne diffèrent entre elles que par la manière de mesurer dans l'espace la position des espèces. Il en est conclu que ces méthodes donnent approximativement un même résultat, pour peu que les structures spatiales sous-jacentes soient suffisamment bien définies sur la zone d'étude.

KEY-WORDS: Plant biogeography – Herbarium records – List of species occurrences – Multivariate statistics – Canonical Correlation Trend Surface Analysis (CCTSA) – Correspondence Analysis (CA) – Discriminant Analysis (DA) – Discriminant Analysis on Eigenvectors of Neighbourhood Operator (DAENO) – Point pattern – Spatial analysis

1. Introduction

The understanding of the geographical zonation of different plant species in a given region is of importance for ecologists because it may help them to develop hypotheses on the ecological history of a region and on the underlying environmental variables driving the observed vegetation patterns (HIRZEL & al., 2003; SPICHIGER & al., 2004). To identify zonation in a given area, many ecological studies focus on the relations between the distributions of several selected species. They try to define a typology grouping together species showing a similar distribution and emphasizing the differences between species showing dissimilar distributions (AUSTIN, 1985; GIMARET-CARPENTIER & al., 2003).

In phytosociology, this issue has been raised frequently (TER BRAAK, 1985; TER BRAAK & LOOMAN, 1986). This field of ecology focuses on processes occurring at a relatively small scale, and the studies of spatial zonation of species is generally based on a systematic sampling of the study area. For example, a virtual grid is superimposed over the study area, and the number of individuals of species sampled in each quadrat is recorded. The statistical methods used to analyse such data are essentially multivariate, as the number of studied species is often large. Ordination methods such as correspondence analysis are generally used in these surveys (HILL, 1974). These methods are optimal when used to show the distribution patterns and vegetation structures in a given area (TER BRAAK & LOOMAN, 1986).

On a larger scale (e.g. continental scale), such designs are difficult to implement, and biologists often look for other sources of information to identify this geographical zonation. In biogeographical studies, existing collections of herbarium specimens may be of great use because the location from where each sample was collected is known (GIMARET-CARPENTIER & al., 2003). In this paper, we use the term "occurrence" to describe such data. According to GIMARET-CARPENTIER (1999), an occurrence corresponds to the location of a species specimen in an area that was gathered by an ecologist in an observational study. In the case of herbarium specimens, an occurrence is therefore characterised by three variables: its X and Y coordinates, and the species to which it belongs.

A list of occurrences is fundamentally different from the data collected in a study using systematic sampling. From a statistical point of view, species occurrences have a particular status. The location of an occurrence at a given point ensures that the species was present there, whereas the absence of an occurrence does not guarantee the absence of the species. An absence may also indicate that during the sampling this species was not recorded at the location (GREEN, 1971; HIRZEL & al., 2003). The spatial distribution of the occurrences reflects the distribution of the species as well as the distribution of the sampling intensity; this distribution is not necessarily uniform across the area (WILLIS & al., 2003). The aim here was to discriminate between species over the sampled area and not to describe the actual distribution of a single species. If it is assumed that the probability of collecting an occurrence at a given point does not vary among the different species considered then the lists of occurrences can be used to identify the geographical variation in species composition. They can be used even if this probability varies spatially.

In this paper we focus on three statistical methods that can be used to emphasise the geographical zonation of species in a given region based on lists of occurrences: (i) *Correspondence Analysis* (HILL, 1974), (ii) *Canonical Correlation Trend Surface Analysis* (WARTENBERG, 1985; GIMARET-CARPENTIER & al., 2003), and (iii) *Discriminant Analysis on Eigenvectors of Neighbourhood Operator* (MÉOT & al, 1993; SPICHIGER & al., 2004). Since these three methods rely on the same mathematical and conceptual bases our aims were to present mathematical connections between them, and to explore the similarity of results using an example dataset. This virtual example simulates a gradient formed by three virtual species *A*, *B*, and *C* in a study area with 50 occurrences per species (Fig. 1).

2. Correspondence Analysis (CA)

Correspondence Analysis (CA) has been widely used in ecology over the last thirty years. This method has been introduced independently by several authors for the analysis of contingency tables built from the systematic sampling of a given area (ROUX & ROUX, 1967; HATHEWAY, 1971; HILL, 1973). Let the contingency table **T** (Q quadrats × S species) containing the number of individuals of each species in each sampled quadrat. CA assigns a set of numerical values to each quadrat and to each species, such that both the discrimination of the species scores between quadrats and the discrimination of the quadrats scores between species is maximum (THIOULOUSE & CHESSEL, 1992). This analysis is optimal to emphasise the spatial organisation of plant species using this kind of sampling scheme (TER BRAAK, 1985).

This method may also be used to analyse the list of species occurrences (Fig. 2A). A grid is superimposed onto the study area, and the occurrences of each species are numbered in each quadrat of the grid. This leads to a contingency table **T**, with Q rows (i.e. Q quadrats of the grid), and S columns (i.e. S species). Correspondence analysis of **T** assigns scores to both species and quadrats such that two quadrats with similar floristic composition have similar scores, and two species with a similar distribution in the quadrats of the grid also have similar scores. By maximizing the variance of these scores, this analysis maximizes the spatial floristic variation in the study area. This property allows the main structures and patterns present in the study area to be highlighted.

Correspondence analysis of the contingency table built from our virtual examples assigns positive scores to quadrats located at the northwest corner of the study area, and negative scores to the quadrats located at the southeast corner. The simulated gradient is therefore described well by these scores. The structures highlighted by the analysis may be interpreted through the distribution of the occurrence scores by species. For example, the occurrences of species *A* are restricted to the negative side of the factorial axis, indicating that this species is distributed in the northwest of our study area. Likewise, species *C* is restricted to the southeast corner of the study area, and species *B* is intermediate. The main drawback of this method is that the user has to choose the quadrat size, and that results obtained may depend on the size chosen. It can be further argued that the use of a grid of quadrats might hide local features of the data, and therefore "oversmooth" the data.

3. Canonical Correlation Trend Surface Analysis (CCTSA)

Discriminant Analysis (DA) is another possibility that can be used to emphasize geographical patterns displayed by the studied species. DA is optimal when the objective is to separate *a priori* groups using measurements for individuals from several variables (GREEN, 1971; MANLY, 1994). This method returns several linear combinations of the descriptive variables which maximises the discrimination between groups. For example, if X_1, X_2, and X_3 are the descriptive variables, the analysis returns a new synthetic variable $Y = a_1 \times X_1 + a_2 \times X_2 + a_3 \times X_3$, and the coefficients a_1, a_2, and a_3 are computed such that the ratio (between-group variance of Y) / (total variance of Y) is maximum. This analysis therefore assigns a numerical score to each occurrence so that the percentage of variation of the scores explained by the factor species is as high as possible.

The geographical zonation of the species in the study area is shown in Figure 1. The distributions of tree species can be worked out using the variables measuring the spatial distribution of their occurrences (Fig. 2B). Polynomial functions of geographical coordinates of occurrences are often used to achieve this, with f.i. variables x, y, x^2, y^2, $x \times y$, x^3, etc., where x and y are the occurrences coordinates (GIMARET-CARPENTIER & al., 2003). Discriminant analysis of species by such polynomial functions has often been used in ecology and is called *Canonical Correlation Trend Surface Analysis* (CCTSA) (GITTINS, 1968; WARTENBERG, 1985; GIMARET-CARPENTIER & al., 2003). The analysis maximizes the distinctions between species according to the spatial distribution of their occurrences. As for many statistical techniques, the discriminant analysis requires a number of occurrences much larger than the number of variables. However, there is no strict rule to choose the number of polynomial functions to be used in this analysis (MANLY, 1994). With sample sizes commonly encountered in the literature, this number rarely exceeds 10 (e.g. GIMARET-CARPENTIER & al., 2003).

CCTSA is mathematically equivalent to the commonly used *Canonical Correspondence Analysis* (CCA) (TER BRAAK, 1986). The difference is that CCA requires a grid of quadrats (in this case the sampling unit is the quadrat, as in CA), whereas this is not necessary in DA (where the sampling unit is the occurrence). CCA is therefore a special form of DA (TER BRAAK & VERDONSCHOT, 1995).

We present an application of CCTSA using a virtual example. As with CA, CCTSA assigns very positive scores to occurrences located at the northwest corner of the study area Figure 1, and very negative scores to the occurrences located at the southeast corner of the area (Fig. 2B)., The analysis clearly emphasises the gradient simulated in the study area. Smoothing methods, such as lowess regression (CLEVELAND & DEVLIN, 1988), may be used to smooth the occurrences scores in the study area; this may give an even clearer picture of the structure. As in CA the distribution of the scores by species reflects the position of the species on the gradient. The histograms reveal that the species *A* is located at the northwestern end of the gradient, the species *C* at the southwestern end, and the species *B* falls between species *A* and *C*.

Thus, CCTSA gives the same conclusions as CA without any need of excessive discretization. For this reason, CCTSA has been recommended by many authors (GITTINS, 1968; WARTENBERG, 1985; GIMARET-CARPENTIER & al., 2003). The main drawback of this method is that the polynomial functions of occurrence coordinates are often strongly correlated (e.g. the x-coordinate is strongly dependent on its square x^2). This is a problem because DA requires that the discriminating variables are not too strongly correlated (MANLY, 1994). In addition, this correlation implies that a large number of polynomial functions have to be included in the analyses to account for a large part of the spatial variation. To circumvent this drawback some authors have proposed the use of orthogonal polynomials in this analysis instead of the classical polynomials (BORCARD & LEGENDRE, 1994).

4. Discriminant Analysis on Eigenvectors of Neighbourhood Operator (DAENO)

An interesting alternative to CCTSA is to replace the set of polynomial functions of geographical coordinates by a set of coordinates which position the occurrences relative to the other occurrences of the pattern. This type of method constitutes the core of *Discriminant Analysis on Eigenvectors of Neighbourhood Operator* (DAENO).

Firstly, a network of neighbouring relationships relating the occurrences of the pattern needs to be computed. Several algorithms are available to compute such networks (e.g. GABRIEL & SOKAL, 1969; PACE & ZOU, 2000). For simplicity here we use Delaunay triangulation (UPTON & FINGLETON, 1985). This network is built from a Voronoi tessellation of the occurrence pattern (UPTON & FINGLETON, 1985). This tessellation is special kind of decomposition of the study area into a set of polygons, determined by distances to the occurrence pattern (Fig. 3A). Each polygon is associated with an occurrence and includes all the places closer to this occurrence than to any other. The Delaunay Triangulation is derived from this tessellation, and connects two occurrences by a line if their Voronoi polygons share a common edge. In the following analyses two occurrences connected by a line on this graph are considered as neighbours.

It is possible to build a matrix describing the network of neighbouring relationships from this graph. Let $\mathbf{V}$ be a matrix with I rows and I columns (where I is the total number of occurrences, all species being pooled). At the intersection of the row i and of the column j, the matrix contains 1 if the i^{th} occurrence is a neighbour of the j^{th} occurrence, and 0 otherwise. The matrix $\mathbf{V}$ is called neighbourhood operator and it can be used to derive a set of scores which describe the position of occurrences.

The diagonal matrix $\mathbf{D}_n$ is computed as following:

$$\mathbf{D}_n = \text{Diag}(\mathbf{V}\ \mathbf{1}_n)$$

where $\mathbf{1}_n$ is the n-vector of 1.

At the intersection of the row i and of the column i, $\mathbf{D}_n$ contains the number of neighbours of the occurrence i. Then, the matrix $\mathbf{S}$ is computed by:

$$\mathbf{S} = \frac{1}{m}\mathbf{D}_n - \frac{1}{m}\mathbf{V}$$

where m is equal to the number of pairs of neighbours (the sum of all values in **V**).

The diagonalization of **S** returns a set of eigenvectors that are orthogonal (i.e. uncorrelated). The first vector assigns a numerical score to each occurrence so that the spatial autocorrelation of the scores is as high as possible (Fig. 3B). In other words, two neighbouring occurrences will have similar scores whereas two very distant occurrences will have very different scores. The second vector maximizes the spatial autocorrelation under the constraint of orthogonality with the first vector, and so on. Theoretical justifications for the above formulas can be found in MÉOT & al. (1993).

Eigenvectors from the diagonalization of **S** can be used in spatial analysis in place of the polynomials of geographical coordinates, such as f.i. in DAENO. This latter analysis has many advantages over CCTSA. Firstly, whereas Cartesian coordinate system (i.e. x, y) gives the position of a point in space relative to an arbitrary point of reference (i.e. the origin of space, with coordinates x = 0 and y = 0), the eigenvectors of neighbourhood operators give the position of an occurrence relative to the whole occurrence pattern (i.e. the studied system). Secondly, two occurrences may be close in geographic space, i.e. with similar Cartesian coordinates, but separated by an impassable boundary (e.g. a mountain). The polynomial functions of geographical coordinates do not take into account this possibility whereas the eigenvectors of the neighbourhood operator do, provided that the neighbouring relationships between occurrences on either side of the boundary are deleted from the matrix **V**. Finally, the eigenvectors of the neighbourhood operator are orthogonal, i.e. uncorrelated. This property important when these variables are used as descriptive variables in a discriminant analysis (MANLY, 1994). MORELLET (1998) noted that the eigenvectors of neighbouring operator explains a larger quantity of spatial variability than polynomial functions of coordinates. For all these reasons, the eigenvectors of the neighbourhood operator have been recommended in spatial analyses by THIOULOUSE & al. (1995).

Practical application of DAENO using virtual species distributions

Using Delaunay triangulation we computed a network of neighbouring relationships. Figure 3A shows the resulting neighbouring relationships between the occurrences. From this graph we derived the neighbourhood operator. The neighbourhood operator was then diagonalized and the 12 first eigenvectors of this matrix were kept for DAENO, although there is no strict rule to decide how many eigenvectors should be kept for further analyses (Fig. 3B). As previously mentioned, the analysis shows the simulated gradient. Very positive scores are seen for occurrences in the northwest corner and very negative scores for the occurrences in the southeast corner (Fig. 2C). Note that the difference between the northwest and the southeast appears even more clearly with this analysis than with CCTSA. Likewise, species *A* and *C* are more strongly discriminated with DAENO.

5. Discussion

The three methods presented above may be used to discriminate species according to their spatial distribution. CCTSA and DAENO are forms of DA. CA is also a form of DA (THIOULOUSE & CHESSEL, 1992) achieved through the superimposing of a grid onto the occurrence pattern. A matrix **W** can be derived, with I rows (the I occurrences) and Q columns (the Q quadrats). At the intersection of the row i and of the column j, this matrix contains 1 if the i^{th} occurrence belongs to the quadrat j, and 0 otherwise. Discriminant analysis of the species by the table **W** returns exactly the same results as CA of the contingency table **T** described above.

These three methods are all discriminant analyses of species by variables measuring the position of occurrences. The only difference is the measure of the spatial position of occurrences; this measure involves, for each method respectively:

- quadrats in CA;
- polynomials of the geographical coordinates in CCTSA;
- eigenvectors of neighbourhood operator in DAENO.

Despite the fact that these variables differ from one method to another the results are similar whatever the method used. All three analyses are likely to identify major patterns in the data. However, theoretical arguments lead us to prefer DAENO, as the eigenvectors of neighbourhood operator have a higher ecological meaning (THIOULOUSE & al., 1995). We used Delaunay triangulation to generate neighbouring relationships but we stress that numerous other methods are available in the literature to build such a network (for a review, see OLLIER, 2005).

As stressed by PALMER (1993), "it is one thing to entrust the validity of equations in the abstract, and yet another to entrust our data to them". A biological application of DAENO is presented in this issue (see SPICHIGER & al., 2005) showing how DAENO may be used to analyse the geographical zonation of several tree species in Paraguay using herbarium specimens. Point patterns are very common in ecology and the need for methods dealing with such data is still considerable (UPTON & FINGLETON, 1985). CA, CCTSA and DAENO may have a role in this context. Despite our theoretical preference for DAENO the other methods are also valid tools which may be preferred because of their wider availability in commercial statistical software. These three methods are freely available (see ade4 package for R software, downloadable at http://cran.r-project.org).

ACKNOWLEDGEMENTS

We would like to thank Michelle Price for correcting the English.

REFERENCES

AUSTIN, M. P. (1985). Continuum concept, ordination methods, and niche theory. *Ann. Rev. Ecol. Syst.* 16: 39-61.

BORCARD, D. & P. LEGENDRE (1994). Environmental control and spatial structure in ecological communities: an example using oribatid mites (Acari, Oribatei). *Environm. Ecol. Statist.* 1: 37-61.

CLEVELAND, W. S. & S. J. DEVLIN (1988). Locally weighted regression: an approach to regression analysis by local fitting. *J. Amer. Statist. Assoc.* 83: 596-610.

GABRIEL, K.R. & R.R. SOKAL (1969). A new statistical approach to geographic variation analysis. *Syst. Biol.* 18: 259-278.

GIMARET-CARPENTIER, C. (1999). *Analyse de la biodiversité à partier d'une liste d'occurrences d'espèces: nouvelles methods d'ordination appliqués à l'étude de l'endémisme dans les Ghâts occidentaux*. Thèse de doctorat. Université de Lyon.

GIMARET-CARPENTIER, C., S. DRAY & J. P. PASCAL (2003). Large-scale biodiversity pattern analyses of the endemic tree flora of the western Ghats (India) using canonical correlation analysis of point data. *Ecography* 26: 429-444.

GITTINS, R. (1968). Trend-surface analysis of ecological data. *J. Ecol.* 56: 845-869.

GREEN, R. H. (1971). A multivariate statistical approach to the Hutchinsonian niche: bivalve molluscs of central Canada. *Ecology* 52: 543-556.

HATHEWAY, W. H. (1971). Contingency-table analysis of rain forest vegetation. *In*: PATIL, G. P., E. C. PIELOU & W. E. WATERS (ed.), *Statistical Ecology. Vol.3: Many species populations, ecosystems and systems analysis*: 271-314 Pennsylvania State University Press.

HILL, M. O. (1973). Reciprocal averaging: an eigenvector method of ordination. *J. Ecol.* 61: 237-249.

HILL, M. O. (1974). Correspondence analysis: A neglected multivariate method. *Appl. Statist.* 23: 340-354.

HIRZEL, A. H., J. HAUSSER, D. CHESSEL & N. PERRIN (2003). Ecological-Niche Factor analysis: How to compute habitat-suitability maps without absence data? *Ecology* 83: 2027-2036.

MANLY, B. F. J. (1994). *Multivariate Statistical Methods. A primer, Second edition*, Chapman & Hall.

MÉOT, A., D. CHESSEL & R. SABATIER. (1993). Opérateurs de voisinage et analyse des données spatio-temporelles. *In*: LEBRETON, J. D. & B. ASSELAIN *Biométrie et Environnement*, 45-71. Masson.

MORELLET, N. (1998). *Des outils biométriques appliqués aux suivis des populations animales: l'exemple des cervidés*. Thèse de doctorat. Université de Lyon.

OLLIER, S. (2005). *Des outils pour l'intégration des contraintes spatiales, temporelles et évolutives en analyse de données écologiques*. Thèse de doctorat. Université de Lyon.

PACE, R. K. & D. ZOU (2000). Closed-form maximum likelihood estimates of nearest neighbour spatial dependence. *Geogr. Analysis* 32: 154-172.

PALMER, M.W. (1993) Putting things in even better order: the advantages of canonical correspondence analysis. *Ecology* 74: 2215-2230.

ROUX, G. & M. ROUX. (1967). A propos de quelques méthodes de classification en phytosociologie. *Rev. Statist. Appl.* 15: 59-72.

SPICHIGER, R., C. CALENGE & B. BISE (2004). The geographical zonation in the Neotropics of tree species characteristic of the Paraguay-Paraná Basin. *J. Biogeogr.* 31: 1489-1501.

SPICHIGER, R., C. CALENGE & B. BISE. (2005). The discriminant analysis of the spatial distribution of plant species occurrences: II. Distribution of major tree communities in Paraguay. *Candollea* 60: ??-??.

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 gradient analysis. *Ecology* 67: 1167-1179.

TER BRAAK, C. J. F. & C. W. N. LOOMAN (1986) Weighted averaging, logistic regression and the Gaussian response model. *Vegetatio* 65: 3-11.

TER BRAAK, C. J. F. & P. F. M. VERDONSCHOT (1995). Canonical correspondence analysis and related multivariate methods in aquatic ecology. *Aquatic Sci.* 57: 255-289.

THIOULOUSE, J., & D. CHESSEL. (1992) A method of reciprocal scaling of species tolerance and sample diversity. *Ecology* 73: 670-680.

THIOULOUSE, J., D. CHESSEL & S. CHAMPELY. (1995). Multivariate analysis of spatial patterns: a unified approach to local and global structures. *Environm. Ecol. Statist.* 2: 1-14.

UPTON, G. & B. FINGLETON. (1985). *Spatial data analysis by example. Vol 1: Point pattern and quantitative data.* John Wiley and Sons.

WARTENBERG, D. (1985) Canonical trend surface analysis: a method for describing geographic pattern. *Syst. Zool.* 34: 259-279.

WILLIS, F., J. MOAT & A. PATON (2003). Defining a role for herbarium data in Red List assessments: a case study of Plectranthus from eastern and southern tropical. *Biodivers. & Conservation* 12: 1537-1552.

Submitted on June 27, 2005
Accepted on July 12, 2005

Addresses of the authors: CCa, DC: Université Claude Bernard Lyon, UMR CNRS 5558, Laboratoire de Biométrie et Biologie Evolutive, 43 Boulevard du 11 novembre 1918, 69622 Villeurbanne Cedex, France. Email (CCa): calenge@biomserv.univ-lyon1.fr

RS, CCh: Conservatoire et Jardin botaniques de la Ville de Genève, Ch. de l'Impératrice, 1, CP 60, CH-1292 Chambésy, Switzerland. Email (RS): rodolphe.spichiger@cjb.ville-ge.ch

Figure captions

Fig. 1. – Distribution of the three virtual species studied in this paper: 50 occurrences were simulated for three species A (●), B (△) and C (+). The three species are distributed along a gradient from the north-west of the area toward the south-West. The polygon defines the study area.

Fig. 2. – Basic principle of the three statistical methods compared. Each method is used to analyse the geographical zonation of the three virtual species A (●), B (△) and C (+). The spatial distribution of the occurrences of these species is displayed on Fig. 1.

A. *Correspondence analysis* (CA). A virtual grid is superposed to the study area, and the occurrences of each species are numbered in each of the Q quadrats. Correspondence analysis is then performed on the resulting table **T**. The quadrat scores are then mapped, and the map is interpreted using the histograms of the occurrence scores for each species.

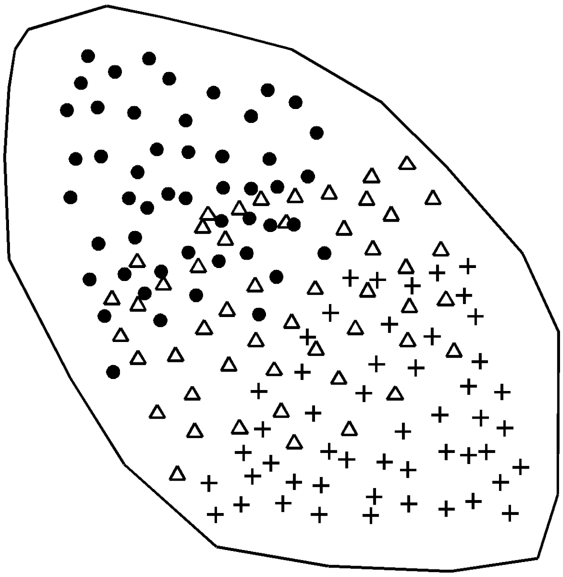
B. *Canonical correlation trend surface analysis* (CCTSA). Polynomial functions of the cartesian coordinates of the species occurrences are in the table **P**, and the species associated with each occurrence is stored in the vector **f**. Discriminant analysis of **P** by the factor **f** assigns scores to each occurrence so that the discrimination between species is maximised according to the spatial distribution of their occurrences. The occurrences scores are then smoothed over the study area using a lowess regression (smoothing is displayed in grey levels). Histograms of the occurrence scores for each species allow the interpretation of the results.

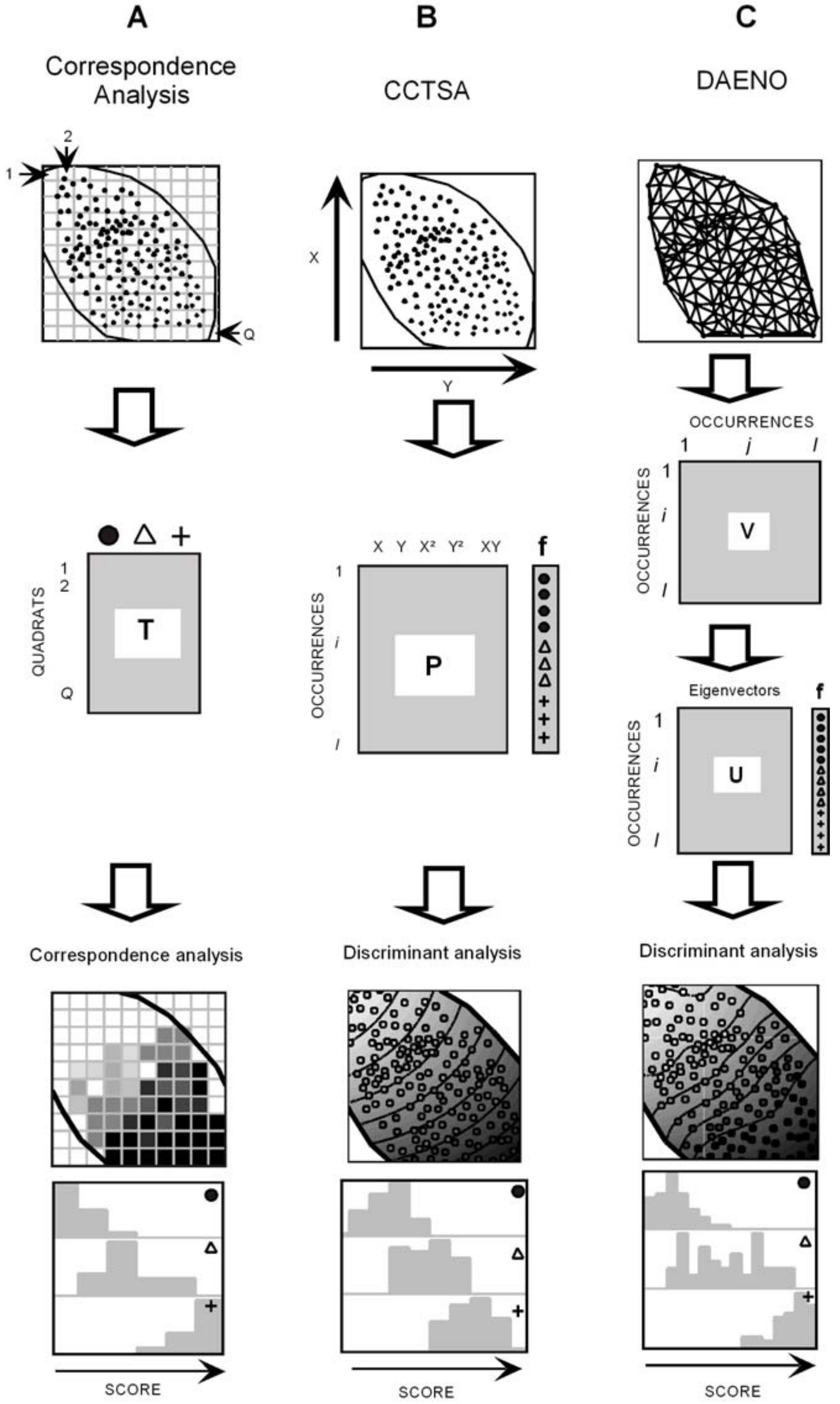
C. *Discriminant analysis on eigenvectors of neighbourhood operator* (DAENO). The neighbourhood operator **V** is a square matrix with a 1 if the occurrence in row and the occurrence in column are neighbours and 0 otherwise. The diagonalization of **V** produces a set of eigenvector, stored in a table **U**, for which the spatial autocorrelation is maximised. A discriminant analysis of **U** by the factor species **f** assigns scores to each occurrence so that the discrimination between species is maximised according to the spatial distribution of their occurrences. The occurrences scores are then smoothed over the study area using a lowess regression (smoothing is displayed in grey levels). Histograms of the occurrence scores for each species allow the interpretation of the results.

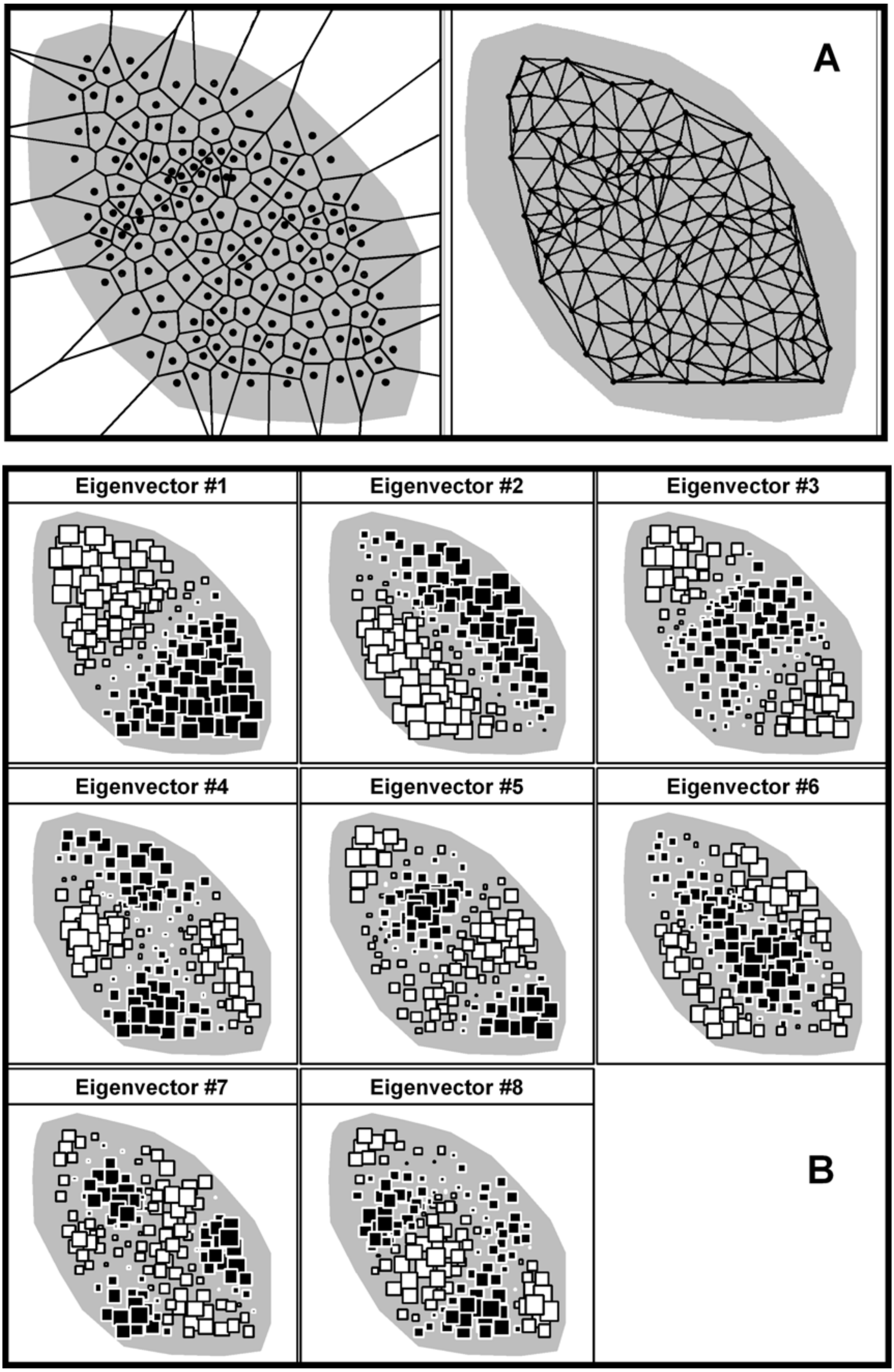
Fig. 3. – Computation of discriminant analysis on eigenvectors of neighbourhood operator.

A. Computation of the network of neighbouring relationships of the occurrences pattern using Delaunay Triangulation (right graph), the dual structure of Voronoï tessellation of the plane derived from the point pattern (left graph). Each line connects two neighbouring occurrences. The neighbourhood operator is computed from this network.

B. Map of the scores given by the first eight eigenvectors of the neighbourhood operator. Each square represents a species occurrence. For a given occurrence, the size of the square is proportional to the absolute value of the score. Black squares correspond to positive values and white squares correspond to negative values. The spatial autocorrelation of the scores is maximized on the first eigenvector. This autocorrelation is also maximised on the second vector, under the constraint of orthogonality with the first one, and so on.




A
Correspondence Analysis
B
CCTSA
C
DAENO
1
2
Q
X
Y
OCCURRENCES
1
j
l
1
i
l
V
QUADRATS
1
2
Q
T
X Y X² Y² XY
f
OCCURRENCES
1
i
l
P
Eigenvectors
f
OCCURRENCES
1
i
l
U
Correspondence analysis
Discriminant analysis
Discriminant analysis
SCORE
SCORE
SCORE


A
Eigenvector #1
Eigenvector #2
Eigenvector #3
Eigenvector #4
Eigenvector #5
Eigenvector #6
Eigenvector #7
Eigenvector #8
B

Annexe 5

Article 5 :

“The discriminant analysis of the spatial distribution of vegetal species occurrences: II. The spatial distribution of major tree communities in Paraguay”

Rodolphe Spichiger, Clément Calenge & Bastian Bise.

Candollea, sous presse

The discriminant analysis of the spatial distribution of vegetal species occurrences: II. The spatial distribution of major tree communities in Paraguay

RODOLPHE SPICHIGER [1]

CLÉMENT CALENGE [2]

&

BASTIAN BISE [1]

[1] *Conservatoire et Jardin botaniques de la Ville de Genève, Ch. de l'Impératrice, 1, CP 60, CH-1292 Chambésy, Genève, Switzerland. Tél. 41 22 418 51 80, Fax (+41) 22 418 51 01, E-mail : rodolphe.spichiger@cjb.ville-ge.ch.*

[2] Correspondence: Clément Calenge, UMR CNRS 5558, Laboratoire de Biométrie et Biologie Evolutive, Université Claude Bernard Lyon 1, 43, Boulevard du 11 novembre 1918, 69622 Villeurbanne Cedex, France. Tel. 33.04.72.43.29.35, Fax. 33.04.72.43.13.88. E-mail: calenge@biomserv.univ-lyon1.fr

ABSTRACT

This paper is the second part in a series of two articles describing theoretical and practical aspects of the spatial discrimination of vegetal species distributions on a given area. We present here a practical case study using herbarium samples. We investigated the geographical zonation of tree species in Paraguay, to identify the vegetal structures and their spatial organisation in this country. Data on the spatial distribution of these species have been collected in the databases of the herbaria of Geneva (Swiss) and of Saint Louis (Missouri, USA). We chose to work on the 32 most common tree species in Paraguay ("ballast" species), to homogenize the sample analysed. We then used a discriminant analysis to identify the main vegetal structures in this country. This study emphasises that the Paraguayan ecotone is built up by several communities, rather than by a continuous change in vegetation composition. We highlighted two main vegetation types in Paraguay, the Chaco and the Parana, themselves made of several "sub-structures". On the extreme west of the Chaco, a first community corresponds to the typical xeromorphic flora with *Aspidosperma quebracho-blanco* as leading species. The centre of the Paraguayan Chaco is occupied by a more psammophilous community, with *Schinopsis quebracho-colorado* as dominant tree species. These two homogeneous vegetation types build the "dry Chaco". The last Chaquean communities highlighted in this study are the xeromesophilous forests of the wet Chaco, characterised by temporarily waterlogged soils and a wetter climate. Two Paranean communities have also been highlighted, on the eastern bank of the Rio Paraguay: (i) the typical Paranean forests, built mainly by *Lauraceae*, and (ii) at the extreme north of the Paraguay, a mixing of Paranean species and of cerrado-related species (savanna species). This mosaic constitutes a transition area between the Paranean forests and the south-Brazilian cerrados. The Chaco and the Parana are separated by an ecotone characterised by a very high diversity, on the eastern bank of the Rio Paraguay. The good agreement of these results with previous results in the literature is a strong argument in favour of the use of herbarium data in biogeographical studies.

RÉSUMÉ

Cet article est le second d'une série de deux articles décrivant les aspects théoriques et pratiques de la discrimination spatiale des distributions de plusieurs espèces sur une zone donnée. Nous présentons ici un cas d'étude pratique utilisant des données d'herbier. Nous avons étudié la zonation géographique des espèces arborées au Paraguay, afin d'identifier les communautés végétales et leur organisation spatiale dans ce pays. Des données sur la distribution spatiale de ces espèces ont été rassemblées dans les bases de données des herbiers de Genève (Suisse) et de Saint Louis (Missouri, USA). Nous avons choisi de travailler sur les 32 espèces les plus communes au Paraguay (espèces “ballast”), pour homogénéiser l'échantillon analysé. Nous avons ensuite utilisé une analyse discriminante pour identifier les principales structures végétales rencontrées dans ce pays. Cette étude souligne que l'écotone paraguayen est caractérisé par plusieurs communautés, et non par un changement continu de la composition de la végétation. Nous avons mis en lumière deux types végétaux principaux au Paraguay, le Chaco et le Parana, eux-mêmes constitués de plusieurs “sous-structures”. A l'extrême ouest du Chaco, une première communauté correspond à la flore xéromorphique typique, dont

Aspidosperma quebracho-blanco est une espèce caractéristique. Le centre du Chaco paraguayen est occupé par une communauté plus psammophile, avec *Schinopsis quebracho-colorado* comme espèce arborée dominante. Ces deux types homogènes de végétation constituent le "Chaco sec". La dernière communauté chaquéenne mise en évidence dans cette étude est formée des forêts xéromésophiles du Chaco humide, caractérisées par des sols temporairement inondés et un climat plus humide. Deux communautés paranéennes ont également été mises en évidence, sur la rive orientale du Rio Paraguay: (i) les forêts paranéennes typiques, principalement constituées par des *Lauraceae*, et (ii) à l'extrême nord du Paraguay, un mélange d'espèces paranéennes et d'espèces liées aux cerrados (espèces de savane). Cette mosaïque constitue une zone de transition entre les forêts paranéennes et les cerrados du Sud-Brésil. Le Chaco et le Parana sont séparés par un ecotone caractérisé par une diversité très importante, sur la rive orientale du Rio Paraguay. L'organisation complexe à petite échelle des communautés végétales souligne l'importance du concept d'échelle dans les études de phytosociologie. L'accord entre ces résultats et ceux de la littérature constitue un argument fort en faveur de l'utilisation des données d'herbiers dans les études de Biogéographie.

KEY-WORDS: Biogeography – Neotropics – Paraguay – Paraná – Chaco – DAENO – floristic communities – scale – herbarium data.

1. Introduction

The study of the geographical zonation of several species distribution ranges is at the very core of Biogeography, because it may help biologists to draw hypotheses on many aspects of the ecology of a region, including environmental and historical aspects. However, at the scale of a country or a continent, it is often hard to design a systematic sampling of the area (e.g. a grid of quadrats covering the whole area). As a result, many biologists try to derive conclusions on this organization using data originating from other sources. In this context, herbarium samples may be of great use, because each sample is an occurrence of the species, i.e. the location of a species specimen on an area, gathered by an ecologist in an observational study (GIMARET-CARPENTIER, 1999). These occurrences carry a certain amount of information on the geographical distribution of the species. In the first paper of this series (CALENGE & al., 2006, same issue), we indicated that several statistical methods exist to analyse such lists of occurrences, and therefore identify the geographical zonation of selected species on a given area. In the present paper, we use one of these methods to analyse herbarium data collected in Paraguay.

Actually, the ecology of Paraguay has been of major concern for botanists, and several studies have described the flora and vegetation of this country (CHODAT & VISCHER, 1916; BERNARDI, 1984; BERNARDI, 1985; LOPEZ & al., 1987; RAMELLA & SPICHIGER, 1989; SPICHIGER & al., 1991; SPICHIGER & al., 1992; SPICHIGER & al., 1995). Therefore, a considerable amount of herbarium samples has been collected by botanists in this area, and most of these specimens are available in the herbarium databases of both the Missouri (Saint Louis) and Geneva Botanical Garden. These data carry a lot of information concerning the spatial organisation of the vegetation in Paraguay, a very complex region from this point of view.

Indeed, the Paraguayan territory is a huge transition area where various vegetation-types, floras and faunas compete together. The biogeographical status of the Paraguay-Paraná basin does not allow a high level of isolation, and therefore of speciation. On the contrary, the large rivers such as the Río Paraná, the Río Pilcomayo or the Río Paraguay and some of their tributaries serve as migration routes for modern floras and faunas (BERNARDI, 1984). Also, altitudinal belts of vegetation are scarce because of the rarity of topographical relieves. Actually, the climatic and edaphic xericity is much higher in the Andean Piedmont than in the surroundings of the Río Paraná, which results in a division of the Paraguayan territory into two biomes separated by the Río Paraguay (fig. 1A). The Chaco extends westward from the bank of the Río Paraguay to the Bolivian boarder, whereas the Paraná area is located at the East of the country, between the Río Paraguay and the Río Parana. The Paraguayan vegetation thus changes from semi-deciduous forests in the Southeast, to xeromorphic forests and thickets in the Northwest.

A recent study of the distribution in South America of tree-species growing in Paraguay highlighted four dispersal centres for the predominant Paraguayan tree species on the continent (SPICHIGER & al., 2004). The Columbian pole and the Paranean pole are characterized by a rather wet climate, whereas the Chaco and the São Francisco area are characterized by a drier climate. Gradients, i.e. continuous changes in vegetation composition, have been highlighted between the Columbian pole and the São Francisco pole ("peri-amazonian" gradient), between the Paraná and the São Francisco pole, and between the Chaco and the three other poles. This study also confirmed the position of the Paraguay territory as an ecotone between the Chaco, the Paraná and the southern Brazilian cerrados ecosystems. A lot of the species growing in the Paraguay-Paraná basin are habitat generalists, i.e. with wide ecological amplitude (OLIVEIRA-FILHO & RATTER, 1995). Using the terminology of MUELLER (1973), the generalist species are polycentric, i.e. related to two, three or four dispersal centres whereas others are monocentric, belonging either to the Paraná or to the Chaco area. However, uncertainty remains concerning the spatial organisation of these communities at the scale of the Paraguay. Indeed, it is now a well-known fact that an ecological process may have different effects at different spatial scales (LEVIN, 1992).

It is a question of importance, because this country is at the crossroads of the four dispersal centres identified in South America. We used here the herbarium samples collected in Paraguay to emphasize the geographical zonation of tree species at the scale of this complex region. This study has two objectives: (i) to determine whether the Paraguayan ecotone is characterized by one or several gradients, with continuous changes of the vegetation composition, or if it is rather built up of several homogenous communities neatly separated; (ii) to determine whether the analysis of herbarium data gives results in agreement with the existing knowledge of the literature, i.e. derived from the analysis of more "traditional" data.

2. Material and methods

We gathered data in the herbaria of the "Conservatoires et Jardins Botaniques de Genève" (Geneva, Switzerland) and of the "Missouri Botanical Garden" (Saint Louis, Missouri, USA). Each herbarium sample is a species occurrence, and is characterized by three variables: its geographical coordinates (longitude and latitude) and the species to which it belongs. However, there are more than 500 tree species in Paraguay, and the herbarium samples of the databases may have been collected for a wide variety of reasons (conservation, systematics, etc.). All these species do not have the same value for the biologists who collect samples, and this value often varies with the context in which samples are collected. The herbaria databases therefore contain very heterogeneous data, and it is likely that the probability of sampling an occurrence at a given place varies between species, which violates the most important hypothesis of the following analyses (CALENGE & al., 2006, same issue).

For this reason, we defined a particular class of tree species, which we termed “ballast species”. This class includes only the most common tree species encountered in Paraguay. We based ourselves on our extensive biological knowledge of the region to select the species belonging to this class (SPICHIGER & RAMELLA, 1998). For example, it includes *Balfourodendron riedelianum*, because it is a common species along the banks of the Paraná and Uruguay river systems, and is also exploited as a commercial timber species. By focusing only on these ballast species, we restricted the analyses to a more homogeneous class of species from the point of view of their sampling probability. Thus, even if there is no definitive means to ensure that the hypothesis of equal sampling probability between species is correct, we will assume it in the rest of this paper on these grounds. Note that the commonness of these species also ensured that the spatial distribution of the occurrences covered approximately the spatial range of the species. Finally, we also required that at least 5 occurrences were present for each species to allow the analysis.

As a result, the database we analysed contained 669 occurrences belonging to 32 species (species names are listed in the Appendix). We analysed the geographical zonation of these tree species using a Discriminant Analysis on Eigenvectors of the Neighbouring Operator (DAENO). The principle of this analysis and the reasons for this choice are detailed in the first paper of this series (CALENGE & al., 2006, same issue). Basically, this method is a discriminant analysis of the species according to synthetic variables describing the spatial position of occurrences on the study area. These synthetic variables are derived from a network of neighbouring relationships relating the occurrences. We used here the Delaunay triangulation algorithm to generate this network (RENKA, 1996; CALENGE & al., 2006, same issue). The DAENO assigns a score to each occurrence so that two species with a similar spatial distribution have a similar average score, and two species with a very different spatial distribution have very different average score. A typology of the species and of the area can then be derived.

Note that the area around Asunción, the capital of the Paraguay, was more extensively sampled than any other part of the country (fig. 1B). We stress here that the non-uniform sampling effort on our study area is not a limit to the use of the method. Indeed, the DAENO discriminates between tree species according to their spatial distribution. Therefore, even if the whole sampling effort varies on the area, the method is still valid provided that the sampling probability is the same for all species at any given point of the study area (CALENGE & al., 2006).

We tested the pertinence of this analysis using a randomisation test. This test was performed by comparing the eigenvalues of the DAENO of the actual data set with the eigenvalues of the DAENO of simulated data sets describing a random distribution of the species in Paraguay. These data sets were generated by randomly permuting the levels of the factor species while keeping the table of eigenvectors of the neighbourhood operator fixed.

All analyses were carried out using the R software (IHAKA & GENTLEMAN, 1996). We used the R packages ade4 and adehabitat for analyses, and the R package tripack for the computation of neighbouring relationships. The R software is freely available on the internet at the URL: http://www.cran.r-project.org/.

3. Results

3.1. Spatial structures

We performed a discriminant analysis on the first 12 eigenvectors of the neighbourhood operator by the factor species. This DAENO produced two highly significant factors ($P < 0.001$). The percentage of geographical variation explained by the species is of 66% and 30% for the first and second factorial axis respectively. This percentage sharply decreases for the next factors (21%, 19% and 17% for the third, fourth and fifth axis respectively), so we restricted our interpretation to the first two factors. The distribution of the species occurrences on the first factor is markedly bimodal (fig. 2A). The lower density of occurrences with a score of about -0.5 indicates the presence of a "frontier" separating two distinct communities. In fact, the Chaquean biome, on the negative side, is opposed to the biome of the southeast of Paraguay (Paraná), on the positive side (fig. 2B). The boundary between these two areas is clearly formed by the Río Paraguay. This strong opposition between the two areas invalidates the hypothesis of a gradient in vegetation composition. The difference between the Chaco (western Paraguay) and the Paraná (eastern Paraguay) is the main vegetation structure in the data, and should not be ignored in the interpretation of the second factor. We therefore interpreted the meaning of the second factor while simultaneously taking into account the strong opposition identified by the first factor, by studying the distribution of the points on the first factorial plane (plane F1-F2, fig. 3).

The occurrences are not homogeneously distributed on this plane, and clusters of points can be identified. These clusters correspond to several communities in the two main biomes. We visually defined a typology of the occurrences according to their position on the first factorial plane (plane F1-F2, fig. 3), to make the interpretation easier. Thus, the tree occurrences have been categorized into six major types: three Chaquean types, labelled A0, A1 and A2, and three Paranean types, labelled B0, B1, and B2. Each type is a community characterized by a rather homogeneous floristic composition (except the group B0, see below), and is geographically well-delimited (fig. 4A). Two points are worth noting concerning the definition of these communities. First, the community B0 seems to be formed by two parallel structures. Actually, this is an artefact caused by the strong sampling effort around Asuncion (Fig. 1B). The lower part of the structure is exclusively made up of these occurrences and the upper part is composed of the rest of the area covered by this group. We therefore grouped these two parts to form a single unit. Moreover, the definition of the communities was not obvious in some cases. Thus, the classification was problematic for a small isolated group of occurrences (emphasized by an arrow on fig. 3A). This group corresponds to a small number of occurrences located at the extreme north of eastern Paraguay. From their position on the factorial plane, it was not evident whether this group should be included in community B1 or B2, or even if a new community (e.g. “B3”) should be defined. Biologically, several arguments led us to assign these occurrences to community B2.

These occurrences were from specimens that had been collected in the extreme north of the eastern Paraguay, an area characterized by a high frequency of cerrados species (pers. field obs.). As the community B2 was mainly characterized by a mixing of such species and of Paranean species (see below), we considered that this community included the unusual group of occurrences.

Note that the shape of the cloud of points on the first factorial plane should not be confused with the classical arch/horseshoe effect (HILL & GAUCH, 1980). Indeed, the arch effect is the result of long gradients in vegetation composition, and is characterised by a homogeneous cloud of points distributed along a U-shape on the factorial plane. In this study, we already proved that the variations in the vegetation composition are not homogeneous: a neat frontier exists between two major kinds of vegetation; this opposition is evident on the histogram of the occurrences scores on the first factor of the analysis (bimodal distribution, fig. 2A), on the cloud of points on the factorial plane (very clustered occurrence pattern, fig. 3A), and from the point of view of the vegetation composition (fig. 4B, see below). All these arguments strengthen the idea that the species are distributed into several distinct communities, and not along homogeneous gradients.

3.2. Floristic composition of the structures

The Chaco and the Paraná areas are clearly separated from the point of view of their floristic composition (fig. 4). Most of the tree species under study are either exclusively Chaquean or exclusively Paranean, but a few are encountered in the two regions, like *Jacaranda cuspidifolia*, *Astronium urundeuva*, or *Calycophyllum multiflorum*. The Paraná and the Chaco are both composed of three communities:

Chaquean communities were labelled A0, A1, and A2, and Paranean communities were labelled B0, B1, and B2 (fig. 4).

Chaquean Flora

The community A2 covers a huge area in geographical space, and the restricted area occupied on the factorial plane indicates that its floristic composition is very homogeneous, i.e. that the floristic composition does not vary spatially. This community is mainly made up of *Capparis retusa, Aspidosperma quebracho-blanco* and *Schinopsis quebracho-colorado*. Moreover, *Capparis speciosa*, *Cercidium praecox*, *Anadenanthera colubrina* var. *cebil*, *Geoffroea striata* and *Tabebuia nodosa* are common species in this community type. Other elements of the Chaquean flora are also encountered in this community type, though to a lesser extent. The community A1 is located in the core of the Chaco. This assemblage is made of the same species encountered in the community A2, but their relative frequencies are different. Thus, *Capparis retusa* and *Aspidosperma quebracho-blanco* are less common in this community. The most frequent species are *Schinopsis quebracho-colorado, Acacia praecox, Cercidium praecox* and *Capparis speciosa.* Finally, the Chaquean community A0 is well-delimited in both geographical space and on the factorial plane. The species composition is very different from the two previously described communities. The most common species are *Schinopsis balansae*, *Acacia caven* and *Diplokeleba floribunda*. Other Chaquean species are also encountered, though to a lesser extent.

Paranean Flora

The type B0 is the ecotone separating the Chaco (western Paraguay) and the Paraná (eastern Paraguay). The frequencies of the two types of species are similar in this area. This assemblage covers a large area on the first factorial plane, indicating that the floristic composition is spatially heterogeneous. The even distribution of the occurrences on the factorial plane revealed the presence of a gradient, which forms the ecotone between the community A0 and B1. Thus, the areas close to the Chaco are richer in Chaquean species, and conversely, the areas adjacent to the Paraná are dominated by Paranean species. The floristic diversity of this area is the highest in the Paraguay. The community B1 occupies the largest part of the eastern Paraguay, and the smallest area covered on the first factorial plane indicates that the floristic composition is spatially homogeneous. Main species encountered include *Nectandra megapotamica*, *Sorocea bonplandii, Balfourodendron riedelianum*, and, to a lesser extent, *Trichilia elegans, Chrysophyllum marginatum* and *Chrysophyllum gonocarpum*. Other species characteristic of the Paraná are also present, though they are less common (*Cochlospermum regium, Patagonula americana, Tabebuia nodosa,* etc.). Finally, community B2 covers a small part of Paranà, i.e. the northeastern part of Paranà. It is characterized by a smaller floristic richness (only 11 species), and by the common presence of *Cochlospermum regium*. Species typical of Paranà are quite frequent in this area, and especially *Nectandra megapotamica, Balfourodendron riedelianum and Sorocea bonplandii*. Again, the small area covered by this community on the factorial plane (fig. 3B) indicates a homogeneous spatial distribution of the species.

4. Discussion

The use of herbarium data led us to identify the main geographical structures in the complex Paraguayan transition area. It is now necessary to compare these results with those of the existing literature, to conclude on the efficiency of herbarium data to identify the vegetation structures in Biogeography. The Paraguayan vegetation does not change continuously from the east to west of Paraguay, but is structured into sharply defined communities. The analysis of the tree occurrences revealed six communities in Paraguay, which compose two main biomes separated by one ecotone. The Chaco is composed of three communities: (i) the typical facies of the xeromorphic forests of the Chaco at the extreme west (A2), (ii) the psammophilous facies with *Schinopsis quebracho-colorado* (A1) and (iii) the wet Chaco, on the west side of the Río Paraguay (A0). The Paraná is composed of two communities: the Paranean semi-deciduous forests (B1), and the forest-cerrados mosaic of the northeastern Paraguay (B2). The Chaco and the Paraná formations are separated by an ecotone (B0), in the Paraguay-Paraná delta (eastern side of the Río Paraguay), where elements of the two main communities are intermingled.

4.1. The Chaco (western bank of the Río Paraguay)

The Chaco is a biogeographical area that has been extensively studied in the literature (for reviews see RAMELLA & SPICHIGER, 1989; SPICHIGER & al., 1991; PRADO, 1993a; PRADO, 1993b), under a wide variety of names: the "Chacoan Province" (TAKHTAJAN, 1986), the "Provincia Chaqueña" (CABRERA & WILLINK, 1973), the "Pantanal and Chaco Phytochorium" (PRANCE, 1989) and the "Regionen des Chaco und seiner Randgebiete" (HUECK, 1966). The Andean uplift during the Oligocene provoked the formation of a synclinal which was first occupied by the sea, leaving, after its withdrawal, salty marine sediments on which Andean deposits accumulated (PUTZER, 1962). This most particular substratum is now colonized by a xeromorphic flora, extending from northwestern Argentina up to northwestern Paraguay and southeastern Bolivia. The Chaco biome is thus likely a Tertiary or early-Pleistocenic refuge although this hypothesis is challenged by MUELLER (1973) who considers the Chaquean biome to be much younger.

Our results agree with the literature (SPICHIGER & al., 1991; SPICHIGER & al., 1995). Indeed, the distribution of the Chaquean species shows two distinct patterns. The communities A1 and A2 correspond to the dry Chaco, also named Boreal Chaco by the Argentinean researchers (MORELLO, 1967; ESKUCHE, 1986). It is centred in the northwestern Chaco, running along the Andean Piedmont and reaching almost to the centre of Argentina. On the other hand, the wet Chaco (community A0) is a transition area corresponding to the drainage of the lower Paraguay and to higher annual precipitations.

The dry Chaco (communities A1 and A2).

The Paraguayan xeromorphic forests and thickets belong to the Chaquean communities of the Gran Chaco (PRADO, 1993b). These communities have been highlighted at the extreme west of the Paraguay in our study (see fig. 4). They are a northern extension of the "Quebrachales" (LEWIS & PIRES, 1981), type of vegetation which has received a large number of names in the literature: "Bosques altos abiertos" *sensu* Argentinian authorities (MORELLO, 1967; ESKUCHE, 1986), "Distrito chaqueño occidental" (CABRERA, 1953; CABRERA, 1976), "Monte" (CÁRDENAS, 1945), "Sukkulenten Dornbusche" or "Monte Alto Formationen" (ESSER, 1982), and "Parque Chaqueño" or "Monte Occidental" (TORTORELLI, 1967). In Paraguay, the Chaquean vegetation and flora are closely associated with loamy or loamy-clayey, alkaline and salty soils which can suffer from either temporary aridity or temporary water logging (SPICHIGER & al., 1991). Actually, the DAENO distinguished two kinds of Chaquean dry forests.

First, the typical facies of "White Quebrachales" (A2) that is characterized by the predominant *Aspidosperma quebracho-blanco*. This medium-sized tree is one of the leading species of the dry Chaquean forest and of its succession stages ("Quebrachales de quebracho-blanco" *sensu* SPICHIGER & al., 1991). Its centre of distribution is the Argentinian and Paraguayan Chaco with extension to the eastern drainage system of the Río Paraguay and to southeast Bolivia. Besides the most common species in this formation, such as *Aspidosperma quebracho-blanco* or *Schinopsis quebracho-colorado,* a lot of smaller trees, spiny shrubs and succulent plants constitute the understorey (*Ruprechtia triflora*, *Cercidium praecox*, *Capparis spp.*). According to SPICHIGER & al. (1991), other species not studied here are also characteristic of this understorey (*Bougainvillea spp.*, *Zizyphus spp.*, *Cactaceae*, *Bromeliaceae*, etc.).

Located at the east of this typical "White Quebrachales" facies, the community A1 corresponds to a more psammophilous forest facies, characterized by the high frequency of *Schinopsis quebracho-colorado* ("Quebrachal de coronillo") and more or less the same understorey (fig. 4B). According to MUNOZ (1990), the "Quebrachal de coronillo" facies and the related thickets are growing in the most xeric areas. Spiny thickets are found in the younger or impoverished stages of both the above-mentioned facies. In the wettest places, species more related to the wet Chaco are encountered (*Tabebuia* *nodosa, Prosopis spp, Acacia spp.*). On the saltiest places ("peladares"), a halophylous steppic vegetation with *Cactaceae, Amaranthaceae* and *Celastraceae* thrives (SPICHIGER & al., 1991).

Some generalist species are also encountered in the dry Chaco (*Astronium urundeuva*, *Maclura tinctoria*, etc., see fig. 4). These species, related to the Paraná-São Francisco gradient and to the Periamazonian gradient (SPICHIGER & al, 2004), invade the dry Chaco in the areas where the substrate is suitable, i.e. on rather well-drained and saltless soils. Thus, gallery-forests allow the extra-zonal mesophilous species (*Astronium urundeuva*, *Maclura tinctoria, Diplokeleba floribunda, Calycophyllum multiflorum* and *Phylostyllon rhamnoides*) to penetrate deeply into the Chaco area (OLIVEIRA-FILHO & RATTER, 1995). Similarly, extra-zonal mesophilous forests with *Anadenanthera colubrina* var. *cebil* and *Astronium urundeuva* grow on the hills and the low montains (Cerro León and Cerro Cabrera). For the mesophilous generalists these hills, serranias and galleries, although they are surrounded by xerophilous forests, act as foreshores between the Paraná area and the Andean Piedmont (SPICHIGER & al., 1991). Finally, cerrado-related species (*Jacaranda cuspidifolia, Astronium fraxinifolium*) grow on top of the fossil sand dunes, serranias and cerros.

The wet Chaco (community A0)

The wet Chaco region is a transition area where many elements are intermingled: the dry Chaco trees, wet savanna species, mesophilous generalists, some scarce Paraná elements, and some anthropogenous pan-American or cosmopolitan species (SPICHIGER & al., 1991; SPICHIGER & al., 1995). It is a southern extension of the Pantanal, which explains the term "Chaco-Pantanal" that has been used for this region. This formation grows on the temporarily waterlogged soils, i.e. on the flooded plains of the Ríos Paraguay and Pilcomayo, and in the Paraná-Paraguay delta. The wet Chaco forests and the palm-savannas form a forest-savanna mosaic landscape. The palm-savannas occupy the part of the land flooded during 3 to 4 months of the year, whereas the forests grow on the higher, well-drained substrates (FIEBRIG & ROJAS, 1933). The forests growing in the Paraguayan wet Chaco are also called "xeromesophilous" forests (SPICHIGER & al., 1991), to distinguish them from the "xerophilous" forests of the dry Chaco. The elements of the xeromesophilous forests are less frost-tolerant than those of the dry Chaco, but are more generalists, growing on temporarily waterlogged basic and loamy soils and on sandier and drier substrates as well. Our study indicates that *Schinopsis balansae*, the "quebracho-colorado", is a leading species of the wet Chaco forests (fig. 4). It also occurs in the Eastern Paraguay community B0. A lot of generalist species (*Astronium urundeuva, Diplokeleba floribunda, Calycophyllum multiflorum, Phyllostylon rhamnoides*) are co-dominant trees together with the Chaquean elements in both the wet Chaco plateau forests and the gallery-forests. In fact, the wet Chaco is also an ecotone (SPICHIGER & al., 1995) between the Paraná area *sensu lato* and the dry Chaco, made up of a mosaic of forest patches intermingled with palm savannas and marshes (see also SPICHIGER & al., 2004). Common savanna tree species, such as *Copernicia alba*, were scarcely represented in the herbaria databases, so we could not take them into account in our analyses. The presence of the marshy areas appears in our analyses through the dominance of *Acacia caven*, and of the gregaríous *Geoffroea decorticans* ("chañares"). The peri-amazonian *Geoffroea spinosa* is also frequent in the wettest places of the dry Chaco, in the wet Chaco and on the eastern bank of the Río

Paraguay. Finally, the wetness of this area, combined with the closeness of the Paraná region, may explain the presence of Paranean species, such as *Tabebuia heptaphylla*.

4.2. Eastern Paraguay (eastern bank of the Río Paraguay)

Eastern Paraguay corresponds to the area delimited by the Paraná and Paraguay rivers. This area is also named Paraguay-Paraná interfluve or Paraguayan Mesopotamy. It was originally covered by a tall semi-deciduous forest, present-day mostly destroyed and replaced by pastures (SPICHIGER & al., 1995). This forest is a piece of the "Misiones Nucleus" which is itself a fragment of the Residual Pleistocenic Dry Seasonally forests (PRADO & GIBBS, 1993). The numerous authors that have worked on this area have given it a wide variety of names: the "Paraná Province" (TAKHTAJAN, 1986), the "Provincia Paranense" (CABRERA & WILLINK, 1973), the "Southern Brazil Phytochorium" (Prance, 1989), or the "Region des Ost- und Südbrasilianischen Wälder" (HUECK, 1966). The Paranean forests spread over the Paraguayan-Upper Paranean-Upper Uruguayan basins (SPICHIGER & al., 1992; SPICHIGER & al., 1995).

The landscape of eastern Paraguay is made of cerros and cordilleras in the centre and the northeast, and of low topographic undulations in other places. The soils are generally clayey in the Paraná drainage, sandier in the Paraguay drainage, rich in nutrients, and frequently hydromorphic in the depressions. Three assemblages are highlighted by our analysis: the Paraná-Paraguay delta (B0); the Paranean semi-deciduous forests (B1); the forest-cerrados mosaic of northeastern Paraguay (B2).

The Paraná-Paraguay delta (community B0)

The Paraná-Paraguay delta ecotone (B0) spreads along the eastern bank of the Río Paraguay and over the Paraguay-Paraná delta. This area is a boundary between the Chaquean and the Paranean flora, and is characterized by very high species diversity. Thus, Chaquean elements as *Schinopsis balansae*, *S. quebracho-colorado*, *Aspidosperma quebracho-blanco*, *Geoffroea decorticans*, *Acacia caven*, *A. praecox*, *Prosopis alba*, *P. nigra*, *Capparis retusa* or *Tabebuia nodosa* invade the eastern bank where they grow in association with Paraná-related species (*Trichilia elegans*, *Chrysophyllum gonocarpum*, *C. marginatum*, *Peltophorum dubium*, *Tabebuia heptaphylla*, *Astronium urundeuva*, *Sorocea bonplandii*, *Cedrela fissilis*, *Balfourodendron riedelianum*, etc.). Generalist species are also encountered there. Field observations indicate that these species grow along the galleries (*Calycophyllum multiflorum*, *Maclura tinctoria*, *Geoffroea spinosa*, etc.) and in the savanna-patches (*Jacaranda cuspidifolia*, *Diplokeleba floribunda*, etc.). Finally, note that *Tabebuia heptaphylla*, together with *Astronium urundeuva* and *Chrysophyllum gonocarpum,* constitute the forest patches in the Paraguay-Paraná delta (ESKUCHE, 1982; SPICHIGER & al., 1992).

The Paraná forest (community B1)

The Paranean species constitute the Paranean semi-deciduous forest ("Floresta Ombrofila Mista" according to Brazilian authors). This forest is centred in the Upper Paranean-Upper Uruguayan basins. It extends from northeastern Argentina to southeastern Brazil and to the southern Planalto, along the Río Paraná and its tributaries. Even if some of them provide a much wider pattern of distribution, a lot of mesophilous species are strongly related to the Paraná area (see fig. 4): *Cedrela fissilis, Balfourodendron riedelianum, Chrysophyllum gonocarpum, Chrysophyllum marginatum, Peltophorum dubium, Tabebuia heptaphylla, Astronium urundeuva, Xylopia aromatica, Patagonula Americana, Sorocea bonplandii* and *Nectandra megapotamica.* Due to the high frequency of *Lauraceae* species*,* these forests are called "Monte de Laurel", "Laurisilva", or "Parapiptadenia rigida-Nectandra saligna" forests by the Argentinian researchers (ESKUCHE, 1982; ESKUCHE, 1984). In Paraguay, the *Lauraceae - Cedrela fissilis - Chrysophyllum gonocarpum* tall forests grow on the well-drained clayey soils and can be considered as the climacical vegetation (SPICHIGER & al., 1992). This type of vegetation seems very homogeneous in our study, as the area covered by the community B1 on the factorial plane is very small compared with the other communities (see fig. 3). Three facies are usually distinguished in the Paranean forests (SPICHIGER & al., 1992): a typical facies with *Balfourodendron riedelianum*; a riparian facies with *Guarea kunthiana* and *Euterpe edulis*; and, a psammophilous facies with *Helietta apiculata*. *Lauraceae* tall forests with *Luehea divaricata* and *Plinia rivularis* are confined to the swampy substrates. We were unable to identify these patterns in our study, as most of these species were not studied here.

The forest-cerrado mosaic (Canendiyu, Concepcion and Amambay areas, community B2).

The analysis pointed out that the community B2, in the northeast of Paraguay, occupies a small area on the first factorial plane (fig. 3). Although at the scale of our study, this community seemed rather homogeneous, at a smaller scale, savannas and tall *Lauraceae* forests outline a mosaic (pers. field obs.). This again stresses the importance of the spatial scale in ecological studies (LEVIN, 1992). This area is an ecotone between the cerrados of the southern Brazil and the Paranean flora, where the above-mentioned Paranean elements are intermingled with cerrado-related species (*Cochlospermum regium, Jacaranda cuspidifolia*). Other species characteristic of the cerrados are found there, though not included in this study (*Pradosia brevipes, Anacardium humile, Andira laurifolia,* and other geoxylic subshrubs *sensu* PENNINGTON, 1990). These species are growing in the patches of cerrados or in the low forests where *Myrtaceae* are predominant: ecoclinal belts between tall forests and savannas, invasive forest-islets surrounded by grasslands, gallery-forests fringes (SPICHIGER & al., 2002).

4.3. Conclusions on the Paraguayan transition area

On a large scale, the flora of Paraguay appears as an ecotone between the cerrados, the xeromorphic Chaquean forests and the semi-deciduous Paranean forests. The changes in vegetation composition between these types appear continuous. However, this study has shown that, at a smaller scale, a more complex structure emerges. Actually, the Paraguayan flora is made of two main vegetation biomes: the Chaco and the Paraná. Between these two forest masses extends a wide transition area, made up of a mosaic of forest-patches intermingled with palm-savannas, cerrados (Brazilian savannas), pastures and cultivated lands (SPICHIGER & al., 1995). This mosaic of vegetation types also constitutes an ecotone, at small scale, where floristic elements of distinct origins are converging to or diverging from (SPICHIGER & al., 1995). The strong heterogeneity emphasized in this study underlines the importance of the concept of scale in ecology (LEVIN, 1992). A study of the spatial distribution of the tree species could be carried out at an even smaller scale, e.g. at the scale of the wet

Chaco (A0), and a new level of heterogeneity would then appear. Indeed, in the Chaco area the generalist species (e.g. *Callycophyllum multiflorum*) that appear associated to the Chaquean communities in this study, only grow in riverine forests, hills and low mountains. This close association would be clear in a study carried out on an area of only several hectares. This relationship between scale and heterogeneity is therefore a fundamental issue in community studies, and should be fully understood to get further insight in the functioning of any ecosystem.

4.4. Conclusion on the use of herbarium data in biogeographical studies

The study of the geographical zonation of the vegetation on the large scale is often problematic, because the large areas involved renders the systematic sampling of the area difficult. Such studies are even more difficult in tropical regions, because of the limited access of the biologists to remote areas. Though herbarium data do not have the same quality as systematic sampling, they carry information on the species spatial distribution. Of course, many questions arise on their representativeness of the studied systems. The samples available in the herbarium databases have been collected in a wide variety of contexts by a large number of biologists. For example, the herbarium of Geneva has been created 175 years ago, and many botanists working everywhere in the world have contributed to its development. These sources are so various that the detection of bias is practically impossible from the data. The bias can be supposed by drawing hypotheses on the usual behaviour of the biologists in the field.

Thus, in tropical regions, because many areas are inaccessible to the biologists, the sampling effort is not spatially uniform. Studies relying on herbarium data have to acknowledge that the distribution of the species occurrences is the result of both the species distribution and the distribution of the sampling intensity, and that the later is often unknown. Therefore, the statistical methods used have to take into account this characteristic of the data. In this study, we only supposed that the variation in sampling intensity is the same for all species, even if the sampling probability varies from one place to another. However, the hypothesis of equal sampling probability among species is still questionable with herbarium data. By choosing a priori a subset of species in the herbarium databases on a criterion of commonness, we certainly homogenized these probabilities. It is likely that a small residual variation in this probability remains, but as noted above, there is no means to measure this variation.

Another possible source of bias arises if botanists who contribute to the herbarium databases tend to collect and label a specimen according to reconceived vegetation types. In this case the later analysis will return results that are the image of the biologist beliefs. However, because herbarium samples are collected by numerous biologists for a long time, and because all the biologists working on a given area do not necessarily have the same vision of its vegetation structures, the pooling of all samples in the herbarium database is a mix of all these visions (though we do not know in what proportions each is present). The analysis of herbarium data may in this case give a model of the system summarizing all these visions. Although results derived from herbarium databases are not by themselves definitive proofs of the validity of a conception of the vegetation organisation, they are one source of information, which can be confronted to other results from the literature. In our study, the agreement was total with previous results. Herbarium data therefore bring additional clues which can help to build a conceptual model of the spatial organisation of the vegetation of the country, as it was the case here for the Paraguay. We think that it is a strong argument in favour of the use of herbarium data in biogeographical studies.

ACKNOWLEDGEMENTS

We would like to thank Daniel Chessel (University of Lyon, France) for his help with the statistical analyses. We are also grateful to Michèle Price for correcting the English.

REFERENCES

BERNARDI, L. (1984). Contribución a la dendrología Paraguaya. Primera parte. *Boissiera* 35: 1-341.

BERNARDI, L. (1985). Contribución a la dendrología Paraguaya. Secunda parte. *Boissiera* 37: 7-294.

CABRERA, A.L. (1953). Esquema fitogeográfico de la República Argentina. *Revista Mus. La Plata* 8: 87-168.

CABRERA, A.L. (1976). *Regiones fitogeográficas Argentinas. Enciclopedia Argentina de Agricultura y Jardinería*, Second Edition, Vol. II. ACME S.A.C.I., Buenos Aires, Argentina.

CABRERA, A.L. & A. WILLINK (1973). *Biogeogragrafía de America Latina*. Segretaría General OEA, Washington DC.

CALENGE, C., R. SPICHIGER, D. CHESSEL, & C. CHATELAIN (2006). The discriminant analysis of the spatial distribution of vegetal species occurrences: I. theoretical aspects. *Candollea.*

CÁRDENAS, M. (1945). Aspecto general de la vegetación de Bolivia. *in:* VERDOORN, F. (ed.), *Plants and plant science in Latin America*: 312-313. The Chronica Botanica Company, Waltham (U.S.A.).

CHODAT, R. & VISCHER, W. (1916). *La végétation du Paraguay*. J. Cramer, Vaduz.

ESKUCHE, U. (1982). Struktur und Wirkungsgefüge eines subtropischen Waldes Südamerikas. Struktur und Dynamik von Wäldern. *In:* DIERSCHKE, H. (ed.), *Berichte der Internationalen Symposien der Internationalen Vereinigung Vegetationskunde*: 49-64. Vaduz.

ESKUCHE, U. (1984). Vegetationsgebiete von Nord und Mittelargentinien. *Phytoceonologia* 12: 185-199.

ESKUCHE, U. (1986). Bericht über die 17. Internationale Pflanzengeographische Exkursion durch Nordargentinien. *In:* ESKUCHE, U. & E. LANDOLT (ed.) *Contribución al conocimiento de flora y vegetación del Norte de Argentina*: 12-177. Veröff. Geobot. Inst. ETH, Stift. Rübel, Zurich (Switzerland).

ESSER, G. (1982). Vegetationsgliederung und Kakteenvegetation von Paraguay. *Trop. Subtrop. Pflanzenwelt* 38: 5-113.

FIEBRIG, C. & T. ROJAS (1933). Ensayo Fitogeográfico sobre el Chaco Boreal. *Revista Jard. Bot. Mus. Hist. Nat. Paraguay* 3: 1-87.

GIMARET-CARPENTIER, C. (1999). Analyse de la biodiversité à partier d'une liste d'occurrences d'espèces: nouvelles methods d'ordination appliqués à l'étude de l'endémisme dans les Ghâts occidentaux. PhD thesis, University of Lyon.

HUECK, K. (1966). *Die Wälder Südamerikas*. Gustav Fischer, Stuttgart.

IHAKA, R. & R. GENTLEMAN (1996). R: A Language for Data Analysis and Graphics. *J. Comp. Graph. Stat.* 5: 299-314.

LEVIN, S. A. (1992). The problem of pattern and scale in Ecology. *Ecology* 73: 1943-1967.

LEWIS, J. P. & E. F. PIRE (1981). *Reseña sobre la vegetatión del Chaco Santafesino. Serie Fitogeográfica n° 18*, Buenos Aires.

LOPEZ, J.A., E.L. LITTLE, G.F. RITZ, J.S. ROMBOLD & W.J. HAHN (1987). *Arboles comunes del Paraguay. Ñande yvyra mata kuera*. Cuerpo de Paz, Paraguay.

MORELLO, J. (1967). Bases para el estudio fitoecológico de los grandes espacios (el Chaco argentino). *Ciencia e Investigación* 23: 252-267.

MUELLER, P. (1973). *The dispersal centres of terrestrial vertebrates in the Neotropical realm: A study in the evolution of the Neotropical biota and its native landscapes*. Junk, La Hague.

MUÑOZ, J. (1990). Anacardiaceae. *In:* SPICHIGER, R. & L. RAMELLA (ed.) *Flora del Paraguay:* 1-88. Conservatoire et Jardins Botaniques de la Ville de Genève, Geneva.

OLIVEIRA-FILHO, A.T. & J.A. RATTER (1995). A study of the origin of central brazilian forests by the analysis of plant species distribution patterns. *Edinburgh J. Bot.* 52: 141-194.

PENNINGTON, T.D. (1990). Sapotaceae. *Flora Neotropica* 52: 1-774.

PRADO, D.E. (1993a). What is the Gran Chaco vegetation in South America? I. A review. Contribution to the study of the flora and vegetation of the Chaco. *Candollea* 48: 145-172.

PRADO, D.E. (1993b). What is the Gran Chaco vegetation in South America? II. A redefinition. Contribution to the study of the flora and vegetation of the Chaco. *Candollea* 48: 615-629.

PRADO, D.E. & P.E. GIBBS (1993). Patterns of species distributions in the dry seasonal forests of South America. *Ann. Miss. Bot. Gard.* 80: 902-927.

PRANCE, G.T. (1989). American tropical forests. *In*: LIETH, H. & M.J.A. WERGER (ed.) *Tropical rain forest ecosystems*: 99-132. Elsevier, Amsterdam.

PUTZER, H. (1962). *Geologie von Paraguay*. Gebüder Bornträger, Berlin-Nikolassee.

RAMELLA, R. & R. SPICHIGER (1989). Interpretación preliminary del medio físico y de la vegetación del Chaco boreal. Contribución al estudio de la flora y de la vegetación del Chaco. *Candollea* 44: 639-680.

RENKA, R.J. (1996). Algorithm 751: TRIPACK: a constrained two-dimensional {Delaunay} triangulation package. *ACM Trans. Math. Soft.* 22: 1-8.

SPICHIGER, R. & RAMELLA, L. (1998). *Flora del Paraguay*. Conservatoire et Jardins Botaniques de la Ville de Genève, Geneva.

SPICHIGER, R., L. RAMELLA, R. PALESE & F. MERELES (1991). Proposición de leyenda para la cartografía de las formaciones vegetales del Chaco paraguayo. Contribución al estudio de la flora y de la vegetación del Chaco. *Candollea* 46: 541-564.

SPICHIGER, R., B.S. BERTONI & P.A. LOIZEAU (1992). The forests of the Paraguayan Alto Paraná. *Candollea* 47: 219-250.

SPICHIGER, R., R. PALESE, A. CHAUTEMS & L. RAMELLA (1995). Origin, affinities and diversity hot spots of the Paraguayan dendroloras. *Candollea* 50: 515-537.

SPICHIGER, R., C. CALENGE & B. BISE (2004). The geographical zonation in the Neotropics of tree species characteristic of the Paraguay-Paraná Basin. *J. Biogeog.* 31: 1489-1501.

TAKHTAJAN, A. (1986). *Floristic regions of the world*. University of California Press.

TORTORELLI, L.A. (1967). Formaciones forestales y maderas del Paraguay. *Bol. Inst. Forest. Latinoamer. Invest.* 24: 3-34.

APPENDIX

The number of occurrences for each species of interest is displayed in the table below, as well as the abbreviations used throughout this paper.

Species name	Abbreviation	Number
Acacia caven (Molina) Molina	Aca.cav	34
Acacia praecox Griseb.	Aca.pra	25
Anadenanthera colubrina var. cebil (Griseb.) Altschul	Ana.col.ceb	8
Aspidosperma quebracho-blanco Schltdl.	Asp.que	28
Astronium fraxinifolium Schott	Ast.fra	9
Astronium urundeuva (Allemão and M. Allemão) Engl.	Ast.uru	10
Balfourodendron riedelianum (Engl.) Engl.	Bal.rie	32
Bulnesia sarmientoi Griseb.	Bul.sar	9
Calycophyllum multiflorum Griseb.	Cal.mul	38
Capparis retusa Griseb.	Cap.ret	37
Capparis speciosa Griseb.	Cap.spe	13
Cedrela fissilis Vell.	Ced.fis	9
Cercidium praecox (Ruiz and Pav.) Harms	Cer.pra	16
Chrysophyllum gonocarpum (Mart. and Eichler) Engl.	Chr.gon	17
Chrysophyllum marginatum (Hook. and Arn.) Radlk.	Chr.mar	36
Cochlospermum regium (Schrank) Pilg.	Coc.reg	13
Diplokeleba floribunda N. E. Br.	Dip.flo	37
Geoffroea decorticans (Hook. and Arn.) Burkart	Geo.dec	9
Geoffroea striata (Willd.) Morong	Geo.str	11

Jacaranda cuspidifolia Mart.	Jac.cus	11
Maclura tinctoria (L.) D. Don ex Steud.	Mac.tin	20
Nectandra megapotamica (Spreng.) Mez	Nec.meg	44
Patagonula americana L.	Pat.ame	21
Peltophorum dubium (Spreng.) Taub.	Pel.dub	11
Prosopis alba Griseb.	Pro.alb	11
Prosopis nigra (Griseb.) Hieron.	Pro.nig	14
Schinopsis balansae Engl.	Sch.bal	35
Schinopsis quebracho-colorado (Schltdl.) Barkley and Mey.	Sch.que	26
Sorocea bonplandii (Baill.) W.C. Burger, Lanj. and Wess. Boer	Sor.bon	27
Tabebuia heptaphylla (Vell.) Toledo	Tab.hep	18
Tabebuia nodosa (Griseb.) Griseb.	Tab.nod	19
Trichilia elegans A. Juss.	Tri.ele	21

FIGURES LEGEND

Fig. 1. – (A) Map of the Paraguay; the main rivers and climatic areas are indicated. (B) Distribution of the 669 tree occurrences in Paraguay.

Fig. 2. – (A) Histogram of the distribution of the tree occurrences on the first factor of the discriminant analysis of eigenvectors of neighbouring operator by the species. (B) map of the distribution of the occurrence scores on the first factorial axis in Paraguay. The scores are smoothed using lowess regression on 167 neighbours. The grey levels reflect variation in occurrence scores (cf. insert).

Fig. 3. – (A) Scatter plot of the occurrences scores on the F1-F2 factorial plane of the discriminant analysis of the eigenvectors of the neighbourhood operator by the species. The arrow indicates the group of occurrences for which the definition of groups was problematic (see text). (B) Typology of the occurrences on the first factorial plane. Six groups of occurrences with a rather homogeneous floristic composition have been visually defined. Each group is identified by a star connecting all the occurrences of the group to its barycentre on the factorial plane.

Fig. 4. – (A) Geographical position of the six groups of occurrences defined in Fig. 3. Each group is identified by a star connecting the occurrences to the geographical barycentre of the group. The contour polygon of each group is also displayed (Light grey polygons correspond to the Chaquean communities, and dark grey polygons indicate the Paranean communities). (B) Species composition of the six groups of occurrences defined in Fig. 3. The importance of a species in a given community is

represented by a black square. The square size is proportional to the percentage of the total number of occurrences of the community that is represented by this species (cf. insert).

Fig. 1.

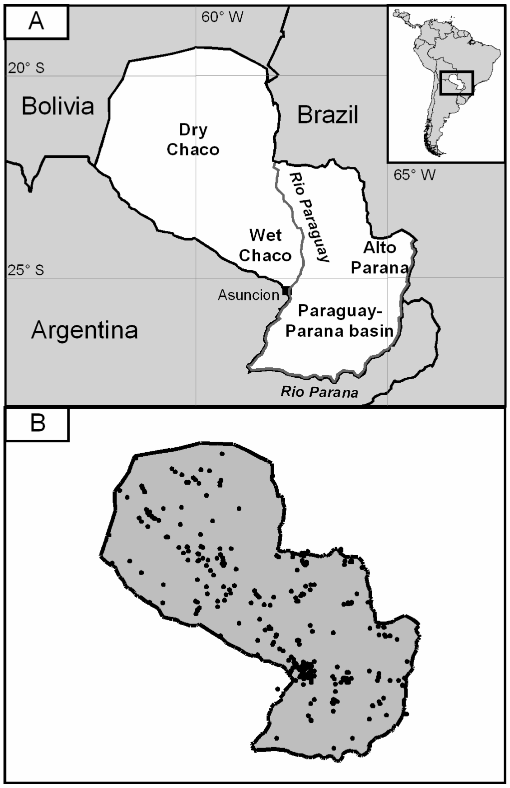

A
60° W
20° S
Bolivia
Brazil
Dry Chaco
Rio Paraguay
65° W
Wet Chaco
Alto Parana
25° S
Asuncion
Paraguay-Parana basin
Argentina
Rio Parana
B

Fig. 2.

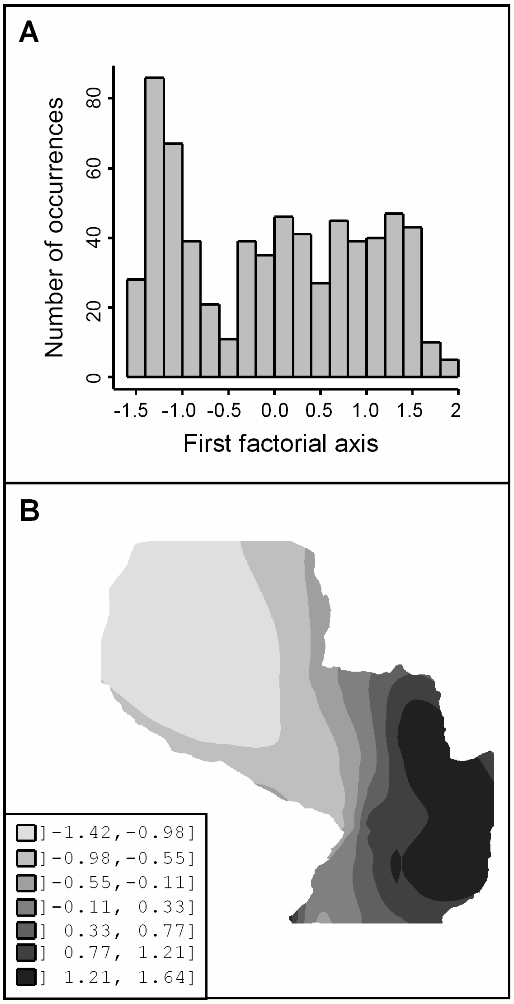

A
Number of occurrences
0
20
40
60
80
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2
First factorial axis
B
]-1.42,-0.98]
]-0.98,-0.55]
]-0.55,-0.11]
]-0.11, 0.33]
] 0.33, 0.77]
] 0.77, 1.21]
] 1.21, 1.64]

Fig. 3.

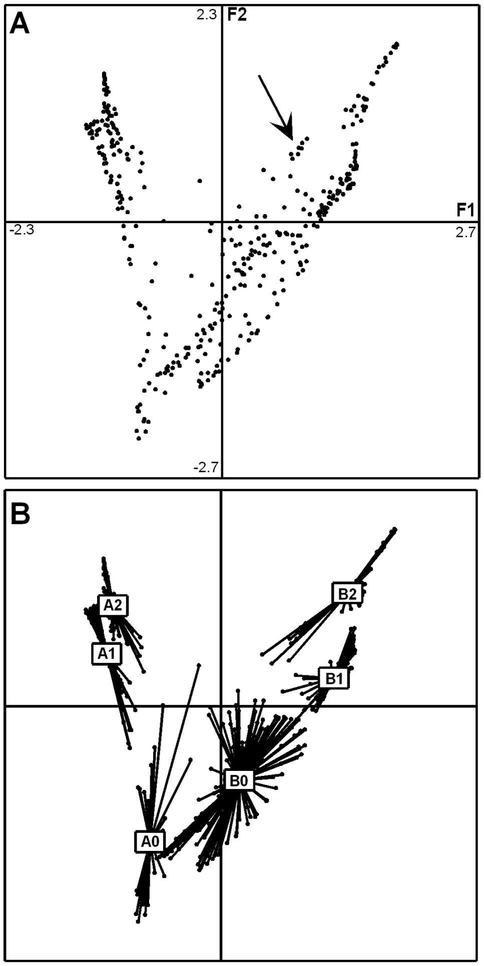
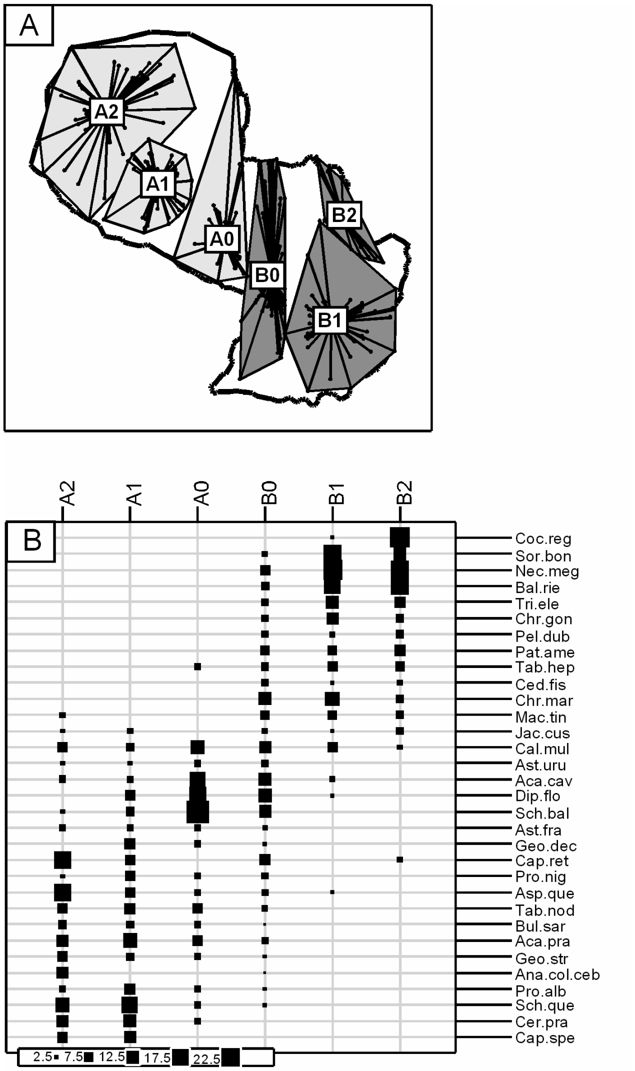


Fig. 4.


A
A2
A1
A0
B0
B2
B1
B
A2
A1
A0
B0
B1
B2
Coc.reg
Sor.bon
Nec.meg
Bal.rie
Tri.ele
Chr.gon
Pel.dub
Pat.ame
Tab.hep
Ced.fis
Chr.mar
Mac.tin
Jac.cus
Cal.mul
Ast.uru
Aca.cav
Dip.flo
Sch.bal
Ast.fra
Geo.dec
Cap.ret
Pro.nig
Asp.que
Tab.nod
Bul.sar
Aca.pra
Geo.str
Ana.col.ceb
Pro.alb
Sch.que
Cer.pra
Cap.spe
2.5
7.5
12.5
17.5
22.5

Annexe 6

Article 6 :

"Biogeography of the forests of the Paraguay-paranà basin"

Rodolphe Spichiger, Bastian Bise, Clément Calenge & Cyrille Chatelain.

Sous presse dans:
T. Pennington (Editeur): "Neotropical Savannas and Dry Forests"

Rodolphe Spichiger, Conservatoire et Jardin botaniques de la Ville de Genève, ch. de l'Impératrice, 1, CP 60, CH-1292 Chambésy, Genève, Switzerland. Tél. 41 22 418 51 80, Fax (+41) 22 418 51 01, (rodolphe.spichiger@cjb.ville-ge.ch)

Bastian Bise, Conservatoire et Jardin botaniques de la Ville de Genève, ch. de l'Impératrice, 1, CP 60, CH-1292 Chambésy, Genève, Switzerland. Tél. (+41) 22 418 51 80, (bastian.bise@cjb.ville-ge.ch)

Clément Calenge, UMR CNRS 5558, Laboratoire de Biométrie et Biologie Evolutive, Université Claude Bernard Lyon 1, 43, Boulevard du 11 novembre 1918, 69622 Villeurbanne Cedex, France. Tel. (+33).04.72.43.29.35, Fax. (+33).04.72.43.13.88. (calenge@biomserv.univ-lyon1.fr)

Cyrille Chatelain, Conservatoire et Jardin botaniques de la Ville de Genève, ch. de l'Impératrice, 1, CP 60, CH-1292 Chambésy, Genève, Switzerland. Tél. (+41) 22 418 51 80, (cyrille.chatelain@cjb.ville-ge.ch)

BIOGEOGRAPHY OF THE FORESTS OF THE PARAGUAY-PARANÁ BASIN

Rodolphe Spichiger, Bastian Bise, Clément Calenge and Cyrille Chatelain

ABSTRACT

Since the Last Glacial Maximum (LGM) the biogeography of Neotropical flora is poorly understood. In this study, two biomes have been investigated: the chacoan and the Paranean biome. Three homogenous communities are identified: the chaco seco, the chaco seco with a psammophilous facies (generally characteristic of sandy soils) and the Paranean semi-deciduous forests. Three ecotonal communities are also defined: the chaco húmedo, the Paranean ecotone intermingled with chaco elements, and the Paranean ecotone with cerrado elements.

In order to describe the affinities of the these communities with other South-American vegetation types we describe three distribution patterns and gradient between the chacoan pole, the Colombian pole, and the Paranean pole.

Integrating Mueller's concept of dispersal centres in a model based on the distribution patterns of selected tree-species we can discuss several hypotheses of floristic history, and especially, the proposition that present-day Paranean forest is a remnant of a residual pleistocenic seasonally dry forest (RPSD flora).

INTRODUCTION

The distribution pattern of the present Paraguayan floras is correlated with specific ecological, edaphic and climatic trends (Spichiger et al., 1995). Indeed, gradients in climatic and edaphic moisture availability from the Río Paraná to the Andean Piedmont explain the division of the Paraguayan territory into two vegetation areas separated in the centre by the Río Paraguay (Figure 1). The chacoan vegetation composed of xeromorphic forests and thickets, extends westward from the bank of the Río Paraguay to the Bolivian border, growing on alkaline soil with loamy texture. Low annual rainfall, 400-1000 mm, characterises this area (Spichiger, 1991). The Paranean semi-deciduous forest located to the east, between the Río Paraguay and the Río Paraná, grows on acid soil with clayey texture. High annual rainfall level, 1500-2000 mm, characterises this area (Olivera-Filho and Fontes, 2000). Furthermore a third vegetation type, the cerrado savanna-like vegetation, is found in northern Paraguay. Its soil is acid like Paranean soil, but its texture is sandy. Low annual rainfall level, 750-1250 mm, characterises this area (Prado, 2003).

The Paraguayan territory can be considered as a type transition area where various vegetation types, floras and faunas meet (Prado, 1993b). Bernardi (1984) shows that the Paraná basin is connected with the Amazon basin, the Andes, and the south of the continent through the large rivers such as the Río Paraná, the Río Pilcomayo and the Río Paraguay and some of their tributaries serve as migration routes for modern floras and faunas. Thus he regards the area as one with a low level of isolation and endemic speciation.

A central issue in biogeography is the identification of variation in vegetation composition in response to climate change. Paleoenvironmental data on the functioning of ecosystems and on the dynamics of the plant communities during the late Quaternary in South America were obtained from paleoclimatic and vegetation studies (for review, see Spichiger et al. 2004). These studies aim to elucidate the processes that have produced the large vegetation formations encountered in South America today.

Two main hypotheses are commonly advanced, based on the present distribution of plant species and on palynological studies (for the review, see Spichiger et al. 2004). According to the first hypothesis, the Wisconsinan period (80,000 -10,000 yrs B.P. corresponding to the last major Ice Age) was characterized by a regression of the forests and a spread of the tropical and subtropical open formations in South America.

The second hypothesis is the residual pleistocenic seasonally dry forest model (Pennington et al., 2000; Prado, 2000; Prado and Gibbs, 1993). According to this, during the Last Glacial Maximum (LGM) seasonally dry tropical forests with an intermingle of rain forest and montane taxa were confined to the wettest regions, in places now occupied by (semi-) evergreen Neotropical forests. The remnants of this once much more extensive mesophilous forest form today a circum-amazonian arc passing through the Paraguay-Paraná-Uruguay basins (Prado and Gibbs, 1993).

The aims of this paper are to analyse the floristic communities of the Paraguay-Paraná basin, based on selected tree-species, to describe their affinities with the other vegetation types in South America and to discuss hypotheses on the floristic history of the region by modelling distribution patterns. For that purpose, we try to answer to the following questions:

1. What the distributions of selected predominant tree-species tell us about the floristic heterogeneity of the Paraguay-Paraná basin?
2. What such distributions tell us about the floristic affinities between the Paraguay-Paraná basin and the other areas of South America?
3. What present-day and modelled distributions tell us about the floristic history of South America in general and the ecotonal status of the Paraguay-Paraná basin in particular?

FLORISTIC HETEROGENEITY OF THE PARAGUAY-PARANÁ BASIN

Method

The Paraguay-Paraná basin is located in the centre of South America, at the confluence of major flora types (Bernardi, 1984). The consequence is a floristic heterogeneity (Spichiger et al., 1995). In this study, we worked on the distribution of 39 common species encountered in Paraguay (Table 1), using data from herbarium specimens as our source. We chose these species, (which are all trees except one hemixyle) because they are predominant and characteristic of the main Paraguayan habitats. Therefore, the geographical distribution of these species should reflect the spatial variation in floristic composition. In the rest of this paper, we will use “occurrence” to mean the attested presence of a species at a specified location.

Because our aim is to discriminate these species according to geographical distribution, we used as spatial discriminant analysis, the Discriminant Analysis of the Eigenvectors of neighbourhood operator (DAENO). This is a multivariate analysis that assigns scores to the occurrences so that the geographical zonation on the area is maximised. The principle of this method is the following: if n is the number of occurrences registred one can calculate from the occurrence pattern the $n \times n$ matrix $\mathbf{V}$, named "neighbourhood operator", indicating the neighbouring relationship between the occurrences. At the intersection of the i^{th} row and of the j^{th} column, this matrix contains 1 if the i^{th} occurrence is a neighbour of the j^{th} occurrence, and 0 otherwise. The neighbourhood relationships were here computed by the Delaunay triangulation (Dale, 1999). Then, a matrix $\mathbf{S}$ is computed::

$$\mathbf{S} = \frac{1}{m}\mathbf{D}_n - \frac{1}{m}\mathbf{V}$$

where m is equal to the number of pairs of neighbours (the sum of all values in $\mathbf{V}$), and $\mathbf{D}_n$ is a diagonal matrix:

$$\mathbf{D}_n = \mathrm{Diag}(\mathbf{V}\ \mathbf{1}_n)$$

with $\mathbf{1}_n$ the n-vector of 1. Therefore, at the intersection of the row i and of the column i, $\mathbf{D}_n$ contains the number of neighbours of the occurrence i. The eigenvectors of $\mathbf{S}$ assign scores to each occurrence so that the score autocorrelation is as high as possible for the study area. These scores can be used to describe the spatial position of each occurrence relative to the others (Thioulouse et al., 1995). Méot et al. (1993) have recommended their use in spatial analyses in place of polynomials functions of geographical coordinates, as they take into account a larger part of spatial variation. The discriminant analysis of these eigenvectors by the factor species has the following properties: (i) the percentage of the spatial variation of the occurrences scores on the first axes explained by the factor species is maximized, (ii) the new axes computed by the analysis are uncorrelated. In other words, two species with a similar spatial distribution have a similar average score, and two species with a very different spatial distribution have very different average score. In this study, we carried out the DAENO at two scales. We first analysed the geographical zonation of the focus species in Paraguay, and then we analysed distributions of the same species at the scale of the whole continent.

Results

Two biomes are clearly separated by the analysis, i.e. the chaco (western Paraguay) and the Paraná (eastern Paraguay) defined by their floristic composition (Figure 2). From the east to the west, the vegetation does not change in a continuously, but is structured in sharply defined communities. A few species are encountered in both areas (e.g. *Jacaranda cuspidifolia* Mart., *Astronium urundeuva* (Allemão) Engl., or *Calycophyllum multiflorum* Griseb.), but most species are exclusively chacoan or paranean.

The DAENO of occurrences reveals six vegetation communities in Paraguay, distributed in the two biomes and separated by one ecotone (Figure 2). The chaco is composed of three communities (Figure 2B): the typical facies of the xeromorphic forests of the chaco at the extreme west (A2), the psammophilous facies (relatated to sandy soils) with *Schinopsis quebracho-colorado* (Schltdl.) Barkley & Mey. (A1) and the wet chaco on the west side of the Río Paraguay (A0). The Paraná is composed of two communities: the Paranean semi-deciduous forests (B1), and the forest-cerrado mosaic of northeastern Paraguay (B2). The chaco and the Paraná formations are separated by an ecotone (B0), which covers a large area on the first factorial plane, indicating that the floristic composition is spatially heterogeneous (Figure 2A). The floristic diversity of this community is the highest in Paraguay, because it includes elements from the two biomes (Figure 2C).

Chacoan Flora

Previous studies have pointed out that the chacoan biome is divided into two groups, the "chaco seco", i.e. dry chaco, and the "chaco húmedo", i.e. wet chaco (Ramella and Spichiger, 1989; Spichiger et al., 1995) (for the review, see Spichiger et al., accepted).

Communities A2 and A1 correspond with the dry chaco. The floristic composition of these two communities is similar, but the relative frequency of the species is different (Figure 2C). *Capparis retusa* Griseb. and *Aspidosperma quebracho-blanco* Schltdl. are less common in community A1. Ramella and Spichiger (1989) and Spichiger et al. (1991) have defined *Schinopsis quebracho-colorado* and *Ruprechtia triflora* Griseb. as representative species of the "chaco seco". *Schinopsis quebracho-colorado* is a very common tree on sandy soils and *Ruprechtia triflora* is a shrub contributing to the understory of the chacoan xeromorphic forests and thickets. The "chaco seco" community is centred in the northwestern chaco, running along the Andean Piedmont and reaching to almost the centre of Argentina. The chacoan vegetation and flora are closely associated with loamy or clayey-loam, alkaline and salty soils which can suffer from either temporary aridity or temporary waterlogging (Spichiger et al., 1991). The Paraguayan Gran Chaco represented by the xeromorphic forests and thickets (Prado, 1993b) corresponds with our chacoan community A2, located at the extreme western part of the Paraguay (Figure 2B). This community is mainly made up of *Capparis retusa*, *Aspidosperma quebracho-blanco* and *Schinopsis quebracho-colorado* (Figure 2C).

Community A0 corresponds to the wet chaco (see Spichiger et al., accepted). Its species composition is very different from that of communities A1 and A2. The community A0 is mainly composed of *Schinopsis balansae* Engl., *Acacia caven* (Molina) Molina, and *Diplokeleba floribunda* N.E.Br., and in addition of the common species of A1 and A2 (*Schinopsis quebracho-colorado* and *Aspidosperma quebracho-blanco*), but these are less frequent here (Figure 2C). It is a transition area where species of dry chaco, wet savanna, mesophilous generalists, scarce Paraná elements, and some pan-American or cosmopolitan species are intermingled (Spichiger et al., 1991, 1995, accepted). This area is a southern extension of the Pantanal, which explains the term “Chaco-Pantanal” that has been used for this region (Prado et al., 1992).

Paranean Flora

Our analysis discrimined two communities for the Paraná area, B1 and B2 (Figure 2B). Community B1 occupies the largest part of eastern Paraguay. Main species encountered include *Nectandra megapotamica* (Spreng.) Mez, *Sorocea bonplandii* (Baill.) W.C.Burger, Lanj. & Wess. Boer, *Balfourodendron riedelianum* (Engl.) Engl., and, to a lesser extent, *Trichilia elegans* A.Juss., *Chrysophyllum marginatum* (Hook. & Arn.) Radlk. and

Chrysophyllum gonocarpum (Mart. & Eichler) Engl, and *Patagonula americana* L. (Spichiger et al, accepted) .

Community B2 occupies a small part in the Alto Paraná area (Figure 2B). But it covers a large area on the first factorial plane, indicating that the floristic composition is spatially heterogeneous (Figure 2A). Only 15 species compose its floristic diversity, (Figure 2C). The common presence of *Cochlospermum regium* and *Jacaranda cuspidifolia*, typical species of cerrado (Spichiger et al., 1991) are intermingled in this area with typical species of Paraná, such as *Nectandra megapotamica, Balfourodendron riedelianum* and *Sorocea bonplandii* (Figure 2C). The large area on the first factorial plane and the common presence of both cerrado and Paranean species indicate that the community B2 is an ecotone between the cerrado of the southern Brazil and the Paranean flora.

FLORISTIC AFFINITIES OF THE PARAGUAY-PARANÁ BASIN WITH OTHER PARTS OF SOUTH AMERICA

Results

On a wider scale, the continental analysis revealed that the Paraguay basin vegetation is distributed along South American vegetation gradients. Each gradient represents to the floristic variation between two extreme points, termed "poles" in the rest of this paper. Four poles have been discriminated by the analysis (Spichiger et al., 2004). Three of them are present in Paraguay: the São Francisco, the Paranean and the chacoan poles (Figure 3). This confirms that Paraguay may be viewed as a huge ecotone in South America, at the intersection of the chacoan, the São Francisco and the Paranean poles (Spichiger et al., 2004).“Every species possesses, or used to possess, at least one dispersal centre that was its centre of origin.

During the evolution of the taxon, however, the centre of origin and the centre of dispersal can become widely separated from each other" (Müller, 1973).

Using Müller's concept of "Dispersal Centre" (Müller, 1973; Spichiger et al. 2004), some of our species are monocentric (related to only one pole or dispersal centre), whereas most of them are polycentric, distributed in several poles (Table 1).

The chacoan pole

The chaco is a biogeographical area that has been extensively studied (Prado, 1993a, 1993b; Ramella and Spichiger, 1989; Spichiger et al. 1991, accepted). It forms the xeromorphic vegetation of the plains of northern Argentina, western Paraguay and south-eastern Bolivia, and the extreme western edge of Mato Grosso do Sul state in Brazil (Prado, 1993a, Spichiger et al. accepted).

The Colombian pole

The Colombian pole and the Paranean pole have many species in common (Spichiger et al., 2004). This may explain the curved shape of the scatter plot on the factorial plane 2-3 (Figure 3). The most characteristic species are *Cedrela fissilis* Vell., *Trichilia elegans, Cabralea canjerana* (Vell.) Mart., and *Maclura tinctoria* (L.) D. Don ex Steud., which are widespread in both (semi-)evergreen forests, including gallery forests, and seasonally dry forests. They are mostly generalists, reaching their southern range in the Paraguay basin

The São Francisco pole

This pole corresponds to the biogeographic caatingas area (Cabrera and Willink, 1973; Prado, 2003). It is located in northeast Brazil, occupying the Rio São Francisco basin. Residual pleistocenic seasonally dry flora elements (Pennington et al., 2000; Prado and Gibbs, 1993), e.g. *Astronium urundeuva, Anadenanthera colubrina* var. *cebil* (Griseb.) Altschul, *Peltophorum dubium* (Spreng.) Taub. and cerrado-related species, e.g. *Astronium fraxinifolium* Schott*, Jacaranda cuspidifolia, Cochlospermum regium, Duguetia furfuracea* (A. St-Hill.) Benth. & Hook. f.*,* are characteristic. The Francisco area has a drier climate than Paranean and Colombian areas, and is occupied by a mosaic of savannas and gallery forests (Eskuche, 1982).

The Paranean pole

Some monocentric species (*Sorocea bonplandii* and *Nectandra megapotamica*) are only present in the Paraná pole. Moreover, many polycentric elements are strongly related to this dispersal centre: *Balfourodendron riedelianum, Chrysophyllum gonocarpum, Chrysophyllum marginatum, Peltophorum dubium, Tabebuia heptaphylla* (Vell.) Toledo*, Astronium urundeuva, Xylopia aromatica* (Lam.) Mart.*, Peltophorum dubium, Patagonula americana*. (Spichiger et al., 2004).

Finally, *Geoffroea spinosa* Jacq. is related to three poles: Colombian, São Francisco and chacoan (the “chaco húmedo”). This species is characterized by one of the widest ecological amplitudes among the analysed key-species (Spichiger et al., 2004).

FLORISTIC HISTORY OF SOUTH AMERICA IN GENERAL, AND THE ECOTONAL STATUS OF THE PARAGUAY-PARANÁ BASIN IN PARTICULAR

It is difficult to draw conclusions about past climates from present-day distributions of plant species. The range of a species may change markedly over a 10,000 year period (Webb, 1992) even for species unable of rapid spread. However, the analysis of present-day distributions of unrelated species may help to evaluate existing hypotheses and formulate new ones (see review in Spichiger et al., 2004). The method is based on the comparison of present species distribution with climatological and palynological data (Behling, 1993, 1995, 1997a, 1997b, 2002; Colinvaux et al., 1996a, 1996b; Colinvaux, 1997; Markgraf, 1991; Markgraf and Bradbury, 1982). Several authors have described modelling techniques to investigate the relationships between tree-species and their environment (Carpenter et al., 1993; Dirnböck and Dullinger, 2004; Guisan and Zimmermann, 2000; Hirzel et al., 2002; Hirzel and Arlettaz, 2003; Kutzbach et al., 1998; Overton et al., 2002; Woodward, 1987). A model predicting the potential distribution of each species according to environmental and climatic variables would allow testing the effect of various climatic scenarios. In this study, we have modelled the past and present potential distribution of some species, and we compare these modelled distributions with actual present-day distributions (Spichiger et al., 2004).

Data collection

We studied the occurrence of 12 tree-species encountered in Paraguay (Table 1). Our choice was based on species that are both commonly found in South America and frequently collected. These trees are well represented in the four major “affinity poles” (São Francisco, Paranean, Colombian and chacoan poles). Some are related to only one “affinity pole” and so are monocentric; others are distributed in several and so are polycentric (Table 1). Their distribution data have been compiled from the literature (Berg, 2001; Golte, 1993; Ireland and

Pennington, 1999; Meyer and Barkley, 1973; Pennington, 1990; Prado, 1991; Rohwer, 1993; Todzia, 1992) and from the Geneva herbarium and the Missouri Botanical Garden (TROPICOS) databases. We digitized the data on a map of South America, using ArcView GIS (ESRI, 1996) and processed them by a Microsoft ACCESS software.

Some authors (Hugget, 1995; Torres et al., 1997) have demonstrated that tree-species distributions in semi-deciduous forest are influenced by temperature, precipitation and altitude. Six climatic maps have been chosen corresponding to these three factors (Table 2). The climatic data, in 0.5° resolution, were sourced from IPCC (2003), IIASA (2003) and IRI (2003). They have been used to determine the potential distributions of the 12 tree-species

Differences between present-day and potential distributions

Methods and Results

The spatial models used in this section are based on the relationship between environmental parameters and the distribution of selected species. They are not based on statistical assumptions and can be applied even on small data sets. The procedure used here is described by Skov & Borchsenius (1997) and Skov (1999). Distribution modelling is a one-step procedure where a potential map is constructed based on climatic parameters. The application developed by Skov (2000) for ArcView Spatial Analyst (ESRI, 1996) produces similarity distribution maps, which give potential distributions based on point-to-point similarity (Carpenter et al., 1993). These distribution models use a point-to-point similarity metric to quantify the similarity between two sites (Carpenter et al., 1993). This point-to-point similarity metric is continuous and varies between 1, when the environment at a given site corresponds completely to a known locality and 0, when the environment at a given site does not correspond to a known locality. It is worth noting that the point-to-point similarity metric values do not represent probability (Skov, 2000). A value of 0.90 indicates that the environmental conditions deviate less than 10% from the known range but do not imply a 90% likelihood of finding the species in a given site.

The continuous surface of the potential distribution has been converted to a binary potential distribution map by choosing a suitable cut-level. In our study, only the cells with a point-to-point similarity metric superior to 0.95 have been considered.

Between potential and present-day distributions, two situations occurred: one of equilibrium and one of disequilibrium. The former was met when potential and present-day distributions match almost perfectly. On the contrast, the latter occurred when the present-day distribution represents a subset of the potential distribution. This situation of disequilibrium could be explained by an "under-collection" of the species but this is unlikely as we selected common species that are frequently collected. We consider that paleo-climatic factors could better explain this discrepancy.

A situation of equilibrium means a low ecological amplitude for the studied species. Indeed all the six climatic factors used in potential distribution restrict this species to its present-day distribution. On the other hand, a situation of disequilibrium means a wide ecological amplitude for the studied species. In fact no climatic factors restrict this species to its present-day distribution, it potentially has the capacity to grow in other places.

Monocentric species in equilibrium situation

The present-day distribution of *Araucaria angustifolia* (Bert.) O. Kuntze and *Nectandra megapotamica*, representative of the Paranean flora, corresponds to their potential distribution, as do *Schinopsis balansae* and *S. quebracho-colorado* representative of chacoan flora (Figure 4). Concerning *Schinopsis species actual and potential distribution, S. balansae* is found eastward and ,S. *quebracho-colorado* westward. This confirm that *S. balansae* is a predominant tree of the "chaco húmedo" on temporarily waterlogged chacoan soils, but also on other sandier and drier substrates (Spichiger et al., 1995), and *S. quebracho-colorado* is representative of the "chaco seco", his potential distribution reach almost to the centre of Argentina.

Polycentric species in disequilibrium situation

Four species are polycentric (Figure 5). They show a potential continuous neotropical distribution (*Cedrela fissilis*, *Maclura tinctoria*) or a potential discontinuous circum-Amazonian distribution (*Astronium urundeuva*, *Geoffroea spinosa*). All of them show a potential distribution much wider than their present-day distribution. With their potential continuous distribution, *Cedrela fissilis* and *Maclura tinctoria* are defined as generalist because they could grow everywhere in neotropical region.

Discussion

All our monocentric species are in a situation of equilibrium and show a low ecological amplitude, which explains their restricted distribution. Zoologists (Brown, 1982; Cracraft, 1985) have explained the distribution of some birds and butterflies by a refuge of Quaternary forest in the Paraná basin. The present-day distribution of *Araucaria angustifolia* and *Nectandra megapotamica* could represent the remnants of this refuge. Another hypothesis agrees with the congruence between potential and present-day distribution for the chacoan species: the chacoan area is accepted as a pleistocenic refuge postulated by Iriondo and Garcia (1993).

On the other hand, all our polycentric species have a present-day disjunct pattern, which can be considered as a remnant of a once much more expanded distribution. The residual pleistocenic seasonally dry forest hypothesis (RPSDf) (Pennington et al., 2000; Prado and Gibbs, 1993), which states that during the Last Glacial Maximum (LGM) (25,000-15,000 yr

BP) a seasonally dry flora, and not savannas, replaced (semi-)evergreen Amazonian forests, implies that the paleo-distributions of the polycentric species are different from the present one. Indeed their potential distribution indicates that these species should have a much wider present-day distribution, considering the six climatic factors used for modelling (Table 2). According to RPSDf hypothesis, we can then postulate that paleo-climatic factors explain mainly the present-day less extended distributions.

Distributions during the Last Glacial Maximum (LGM)

Methods and Results

Pleistocenic drier periods have been postulated in particular by Absy et al. (1991), Behling (2001) and Van der Hammen & Absy (1994). More precisely the second author suggests that the completely different vegetation type at the LGM reflects a drier and colder climate. His study established that temperature was approximately 5°C lower, and the precipitation, 30% lower. He based his hypothesis on pollen and charcoal records analyzed in organic sediments, in southeastern Brazil.

Following this hypothesis, we modelled the species distribution for the rainfall and temperature data during the LGM. Our first step has been to create climatic layers on the basis of present-day data (mean precipitation per month , mean minimum temperature from IPCC (2003), using a request in the ArcView software (ESRI, 1996). A range of 5° C and 30% of precipitation have been cut off from each cell of these climatic layers.

A similarity model (Skov and Borchsenius, 1997) has been applied to the new layers. The use of this rather crude model indicated three regions (Figure 6). The first corresponds with the past potential distributions of the polycentric species (*Astronium urundeuva*, *Cedrela fissilis*, *Duguetia furfuracea*, *Geoffroea spinosa*, *Maclura tinctoria*) and the Paranean species (*Araucaria angustifolia*, *Nectandra megapotamica*). The second is a compilation of past potential distributions of the chacoan species (*Aspidosperma quebracho-blanco, Schinopsis balansae and S. quebracho-colorado*). A third region illustrates the Paranean hypothesis (Markgraf, 1991; Pennington et al., 2000), i.e. Patagonian steppic and montane elements invading the area occupied today by the Paranean forest.

Discussion

Our modelled patterns corroborate the RPSD flora hypothesis (Pennington et al., 2000; Prado and Gibbs, 1993), which suggests that generalist elements (e.g. *Cedrela fissilis*, *Maclura tinctoria*), cerrado elements (e.g. *Astronium urundeuva*, *Duguetia furfuracea*) and Paranean species (*Araucaria angustifolia*, *Nectandra megapotamica*) have invaded the Amazon basin during the LGM at the expense of the Amazonian elements. To understand the limiting factor of the monocentric Paranean species invasion, we modelled their past potential distribution based only on temperature and withdrawing one degree at the time. Their distribution pattern moves northward when the temperature falls 3°. We may also suppose that the Paraná basin was occupied by steppic and montane elements during a period 3°C cooler than today (see also Behling, 1995; Colinvaux and De Oliveira, 2001; Colinvaux et al., 1996a, 2000; Ledru, 1993). Furthermore a floristic study by Oliveira-Filho (2000) shows that montane taxa (*Clethra*, *Clusia*, *Drymis*, *Hedyosmum*, *Podocarpus*, *Prunus*, *Weinmannia*) are present today in Paraná area.

The past potential distributions of chacoan species (*Aspidosperma quebracho-blanco*, *Schinopsis balansae* and *S. quebracho-colorado*) show a slight spread northward. Indeed these species are drought and frost-tolerant (Pennington et al., 2000) and so are scarcely sensitive to LGM climatic changes. Considering their adaptation to salty soils (Spichiger et al., 1991), they could have remained in the north of chaco during the LGM, where they are still present today.

CONCLUSION

The distribution patterns of the present Paraguayan floras as explained in the introduction are correlated with specific ecological, edaphic and climatic trends (Spichiger et al., 1995), which are summarized in figure 7. This triangle represents the Paraguay-Paraná basin, which is a huge ecotone where São Francisco, peri-Amazonian generalists, chacoan, and Paranean species compete. The A pole is related to alkaline, salty, loamy soils and low rainfall ("chaco seco"); the B to acid, sandy soils and also low rainfall (RPSD flora); the C to clayey or sandy-clayey, acid soils and high rainfall (Paraná). The "chaco húmedo" is found in the middle of the axis formed by pole A-C.

The monocentric species (*Araucaria angustifolia, Aspidosperma quebracho-blanco, Nectandra megapotamica, Schinopsis balansae* and *S. quebracho-colorado*) are present along the axis from pole A to C. They are characteristic of chacoan and Paranean poles. Due to their narrow ecological amplitudes and their small differences between potential and actual distribution patterns, they can be considered as endemic to the region, or at least as the oldest components of the Paraguay-Paraná ecotone. Salty soil and high rainfall are the two factors that probably best explain the past and present-day distribution of respectively chacoan and Paranean species.

The central position of the polycentric species in the triangle (Figure 7) reflects their opportunist behaviour due to their wide ecological plasticity. The modelled distributions, and ecological amplitudes, of the polycentric generalist species growing in the Paraguay-Paraná ecotone (*Cedrela fissilis*, *Maclura tinctoria*) show that they should be able to colonize a wider range, corroborating the RPDS flora hypothesis of Pennington and Prado (Pennington et al., 2000). Furthermore, according to the hypotheses of drier climate during the LGM (Absy et al., 1991; Behling, 2001; van der Hammen and Absy, 1994), the low rainfall of the pole B (Spichiger et al., 1995) agrees with a spread of the characteristic species of RPSD flora we studied (*Astronium urundeuva* and *Geoffroea spinosa*).

During the LGM the Paraná flora, perhaps originating from the Atlantic area, may have invaded the Amazon Basin through the São Francisco pole and the gallery-forests of the cerrado. Furthermore, the South-Amazonian arch pattern of distribution (e.g. *Anadenanthera colubrina* var. *cebil*) shows that bridges exist – or have existed - between the Paraná Basin and the so-called Tucumano-Bolivian forests of the Piedmont area. We can infer that extra-zonal semi-deciduous forests in the chaco (gallery-forests and montane forests) connect and have connected the Andean Piedmont with the Paraná Basin. These connections reinforce the hypothesis of Oliveira-Filho and Ratter (1995), which supports the spread of Paranean semi-deciduous forests by penetrating into the cerrado Province via the basalt-derived fertile soils of western Minas Gerais and southern Goías. The ecotone B2 (Fig. 2a) between Paranean elements and typical species of cerrado agree with the results of Oliveira-Filho and Fontes (2000), which shows that Paranean flora and cerrado have strong links. These authors indicate that cerrados share a much larger proportion (55% of their total species) with the Atlantic semi-deciduous forests than with the Amazonian forests (20%). Furthermore a study has demonstrated the coexistence of montane and Paranean forest elements within a part of the present-day cerrado region between 17 000 and 13 000 years BP (Ledru, 1993). All these lines of evidence confirm our model, i.e. that the Paranean flora invaded the Amazonian basin at LGM through cerrado (Figure 6).

REFERENCE LIST

Absy, M.L. et al., Mise en évidence de quatre phases d'ouverture de la forêt dense dans le sud-est de l'Amazonie au cours des 60000 dernières années. Première comparaison avec d'autres régions tropicales, *Compt. Rend. Acad. Sci. Paris, Sér. Gén. Vie Sci.*, 312, 673, 1991.

Behling, H., Untersuchungen zur spätpleistozänen und holozänen Vegetations- und Klimageschichte der tropischen Küstenwälder und der Araukarienwälder in Santa Catarina (Südbrasilien), *Diss. Bot.*, 206, 1, 1993.

Behling, H., A high resolution Holocene pollen record from Lago do Pires, SE Brazil : vegetation, climate and fire history, *J. Paleolimnol.*, 14, 253, 1995.

Behling, H., Late quaternary vegetation, climate and fire history from the tropical mountain region of Morro de Itapeva, SE Brazil, *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, 129, 407, 1997a.

Behling, H., Late Quaternary vegetation, climate and fire history in the Araucaria forest and campos region from Serra Campos Gerais (Paraná), S Brazil, *Rev. Palaeobot. Palynol.*, 97, 109, 1997b.

Behling, H., South and southeast Brazilian grassland during late quaternary times: a synthesis, *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, 177, 19, 2001.

Behling, H., Carbon storage increases by major forest ecosystems in tropical south America since the Last Glacial Maximum and the early Holocene, *Global and Planetary Change*, 33, 107, 2002.

Berg, C.C., Moreae, Artocarpeae, and Dorstenia (Moraceae). *Flora Neotropica* 83, The New York Botanical Garden, New York, 2001.

Bernardi, L., Contribución a la dendrología Paraguaya. Primera parte. Apocynaceae – Bombacaeceae – Euphorbiaceae – Flacourtiaceae – Mimosoideae – Caesalpinioideae – Papilionatae, *Boissiera*, 35, 341, 1984.

Brown, K.S., Paleoecology and regional patterns of evolution in forest butterflies, in *Biological Diversification in the Tropics*, Prance, G.T., Ed., Columbia University Press, New York, 1982, 255.

Bush, M.B., Late Pleistocene temperature depression and vegetation change in Ecuadorian Amazon, *Quatern. Res.*, 34, 330, 1990.

Cabrera, A.L. and Willink, A., *Biogeogragrafía de América Latina*. Secretaria General OEA, Washington DC, 1973.

Carpenter, G., Gillison, A.N. and Winter, J., DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals, *Biodiversity Conservation*, 2, 667, 1993.

Colinvaux, P.A., *The Ice-Age Amazon and the problem of diversity*, NOW/Huygenslezing, The Hague, the Netherland, 1997.

Colinvaux, P.A. et al., A long pollen record from lowland Amazonia : forest and cooling in glacial times, *Science*, 274, 85, 1996a.

Colinvaux, P.A. et al., Temperature depression in the lowland tropics in glacial times, *Climatic Change*, 32, 19, 1996b.

Colinvaux, P.A., De Oliveira, P.E. and Bush, M.B., Amazonian and neotropical plant communities on glacial time-scales: The failure of the aridity and refuge hypotheses, *Quaternary Science Reviews*, 19, 141, 2000.

Colinvaux, P.A. and De Oliveira, P.E, Amazon plant diversity and climate through the Cenozoic, *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, 166, 51, 2001.

Cracraft, J., Historical biogeography and patterns of differentiation within the South American avifauna: Areas of endemism, in *Neotropical Ornithology. Ornithological Monographs*, 36, Buckley, P.A. et al., Ed., American Ornithologists' Union, Washington D.C., 1985, 49.

Dale, M.R.T., *Spatial pattern analysis in plant ecology*, Cambridge University Press, Cambridge, 1999.

Dirnböck, T. and Dullinger, S., Habitat distribution models, spatial autocorrelation, functional traits and dispersal capacity of alpine plant species, *Journal of Vegetation Science*, 15, 77, 2004.

Eskuche, U., Struktur und Wirkungsgefüge eines subtropischen Waldes Südamerikas, in *Ber. Int. Symp. Int. Vereinigung Vegetationsk.*, Struktur und Dynamik von Wäldern, Dierschke, H., Ed., J. Cramer, Vaduz, 1982, 49.

ESRI, *Using ArcView GIS. The geographic information system for everyone*, Environmental Systems Research Institute, Inc., Redlands, 1996.

Golte, W., *Araucaria: Verbreitung und standortansprüche einer coniferengattung in vergleichender sicht*, Franz Steiner, Stuttgart, 1993, 20.

Guisan, A. and Zimmermann, N.E., Predictive habitat distribution models in ecology, *Ecol. Modelling*, 135, 147, 2000.

Hirzel, A.H.et al., Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data?, *Ecology*, 83, 2027, 2002.

Hirzel, A.H. and Arlettaz, R., Modeling habitat suitability for complex species distributions by environmental-distance geometric, *Environm. Managem.*, 32, 614, 2003.

Hugget, R.J., *Geoecology, an evolutionary approach*, Routledge, London, 1995.

Intergovernmental Panel on Climate Change, *The IPCC Data Distribution Centre,* http://ipcc-ddc.cru.uea.ac.uk/ , 2003.

International Institute for Applied System Analyses, *IIASA Science for Global Insight*, http://www.iiasa.ac.at/, 2003.

International Research Institute for Climate Prediction, *IRI Linking science to society*, http://iri.columbia.edu/, 2003.

Ireland, H. and Pennington, R.T., A revision of Geoffroea (Leguminosae-Papilionoideae), *Edinburgh J. Bot.*, 56, 329, 1999.

Iriondo, M.H. and Garcia, N.O., Climatic variations in the Argentine plains during the last 18000 years, *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, 101, 209, 1993.

Kutzbach, J. et al., Climate and biome simulations for the past 21 000 years, *Quaternary Science Review*, 17, 473, 1998.

Ledru, M.P., Late Quaternary environmental and climatic changes in Central Brazil, *Quatern. Res.*, 39, 90, 1993.

Liu, K.B. and Colinvaux, P.A., Forest changes in the Amazon during the Last Glacial Maximum, *Nature*, 318, 556, 1985.

Markgraf, V., Younger Dryas in southern South America?, *Boreas (Oslo)*, 20, 63, 1991.

Markgraf, V. and Bradbury, J.P., Holocene climatic history of South America, *Striae*, 16, 40, 1982.

Méot, A., Chessel, D. and Sabatier, R., Opérateurs de voisinage et analyse des données spatio-temporelles, in *Biométrie et environnement*, Lebreton, J.D. and Asselain, B., Ed., Masson, Paris, 1993, 45.

Meyer, T. and Barkley, F.A., Revision del genero Schinopsis (Anacardiaceae), *Lilloa*, 33, 207, 1973.

Müller, P., *The dispersal centres of terrestrial vertebrates in the Neotropical realm: A study in the evolution of the Neotropical biota and its native landscapes*, Junk, La Hague, 1973.

Oliveira-Filho, A.T. and Fontes, A.L., Patterns of floristic differentiation among Atlantic forests in Southeastern Brazil and the influence of climate, *Biotropica*, 32, 793, 2000.

Oliveira-Filho, A.T. and Ratter, J.A., A study of the origin of central Brazilian forests by the analysis of plant species distribution patterns, *Edinburgh J. Bot.*, 52, 141, 1995.

Overton, J.McC., Leathwick J.R., Stephens R.T.T. and Lehmann A., Information pyramids for informed ecosystem management, biodiversity and conservation, *Biodiversity and Conservation*, 11, 2093, 2002.

Pennington, T. D., Sapotaceae. *Flora Neotropica*, 52, The New York Botanical Garden, New York, 1990.

Pennington, T.D., Prado, D.E. and Pendry C. A., Neotropical seasonally dry forests and Quaternary vegetation changes, *J. Biogeogr.*, 27, 261, 2000.

Prado, D.E., *A critical evaluation of the floristic links between Chaco and Caatingas vegetation in South America*, Ph. D. Thesis, University of St. Andrews, U.K., 1991.

Prado, D.E. et al., The Chaco-Pantanal transition in southern Mato Grosso, Brazil, In *Nature and Dynamics of Forest-Savanna Boundaries*, Furley, P.A., Proctor, J. and Ratter, J.A., Ed., Chapman and Hall, London, 1992, 451.

Prado, D.E., What is the Gran Chaco vegetation in South America? I. A review. Contribution to the study of the flora and vegetation of the Chaco, *Candollea*, 48, 145, 1993a.

Prado, D.E., What is the Gran Chaco vegetation in South America? II. A redefinition. Contribution to the study of the flora and vegetation of the Chaco, *Candollea*, 48, 615, 1993b.

Prado, D.E., Seasonally dry forests of tropical South America: from forgotten ecosystems to a new phytogeographic unit, *Edinburgh J. Bot.*, 57, 437, 2000.

Prado, D.E., As Caatingas da Americá do Sul, in *Ecologia e conservação da Caatinga*, Leal, I.R., Tabarelli, M. and Cardoso da Silva, J.M., Ed., Universidade Federal de Pernambuco, Recife, 2003, 3.

Prado, D.E. and Gibbs, P.E., Patterns of species distributions in the dry seasonal forests of South America, *Ann. Missouri Bot. Gard.*, 80, 902, 1993.

Ramella, R. and Spichiger, R., Interpretación preliminar del medio físico y de la vegetación del Chaco boreal. Contribución al estudio de la flora y de la vegetación del Chaco, *Candollea,* 44, 639, 1989.

Rohwer, J.G., Lauraceae: Nectandra. *Flora Neotropica* 60, The New York Botanical Garden, New York, 1993.

Skov, F., Spatial modelling for tropical biodiversity assessment, in *Scan GIS'99-Proceedings from the 7th Scandinavian Research conference on Geographical Information Science*, Stubkjaer E. and Hansen, H.S., Ed., Aalborg Universitetsforlag, Aalborg, 1999, 203.

Skov, F., Potential plant distribution mapping based on climatic similarity, *Taxon*, 49, 503, 2000.

Skov, F. and Borchsenius, F., Predicting plant species distribution patterns using simple climatic parameters: a case study of Ecuadorian palms, *Ecography*, 20, 347, 1997.

Spichiger, R. et al., Proposición de leyenda para la cartografía de las formaciones vegetales del Chaco paraguayo. Contribución al estudio de la flora y de la vegetación del Chaco, *Candollea*, 46, 541, 1991.

Spichiger, R. et al., Origin, affinities and diversity hot spots of the Paraguayan dendrofloras, *Candollea*, 50, 515, 1995.

Spichiger, R., Calenge, C. and Bise B., The geographical zonation in the Neotropics of three-species characteristic of the Paraguay-Paraná Basin. *J. Biogeogr.*, 31, 1489, 2004.

Spichiger, R., Calenge, C. and Bise, B., Spatial distribution of tree major tree communities in the Paraguayan transition area, accepted in *Candollea*.

Thioulouse, J., Chessel, D. and Champely. S., Multivariate analysis of spatial patterns: a unified approach to local and global structures. *Environ. Ecol. Stat.*, 2, 1, 1995

Torres, R.B., Martins, F.R. and Gouvea, L.S.K, Climate, soil, and tree flora relationships in forests in the state of São Paulo, southeastern Brazil, *Revta. Brasil. Bot.*, 20, 41, 1997.

Totzia, C.A., A re-evalution of the genus Phyllostyllon (ULMACEAE), *Sida*, 15, 264, 1992.

van der Hammen, T. and Absy, M.L., Amazonia during the last glacial. *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, 109, 247, 1994.

Webb, T.III., Past changes in vegetation and climate: lessons for the future, in *Global warming and biological diversity*, Peters, R.L. and Lovejoy, T.E., Ed., Yale University Press, New Haven, 1992, 59.

Woodward, F.I., *Climate and plant distribution*, Cambridge university press, London, 1987.

Table legends

Table 1 List of the 39 tree species used in the study. The first 12 were used for construction of the similarity distribution model.

Table 2 List of the 6 quantitative climate variables used in the study

Table 1

Species name	Abbreviation	No of records in Paraguay	No of records in South America	total No of records	Major affinity poles	Polycentric/ Monocentric
Anadenanthera colubrina var. cebil (Griseb.) Altschul	Ana.col.ceb	8	115	123	Colombian to São Francisco poles	Polycentric
Araucaria angustifolia (Bert.) O. Kuntze				418	Paranean pole	Monocentric
Aspidosperma quebracho-blanco Schltdl.	Asp.que	28	0	28	Chacoan pole	Monocentric

Astronium urundeuva (Allemão & M. Allemão) Engl.	Ast.uru	10	56	66	São Francisco to chacoan poles	Polycentric
Cedrela fissilis Vell.	Ced.fis	9	88	97	Colombian to Paranean poles	Polycentric
Duguetia furfuracea (A. St-Hill.) Benth. & Hook. f.	Dug. fur	0	88	88	São Francisco pole	Monocentric
Geoffroea spinosa Jacq.	Geo.spr	11	41	52	Colombian to São Francisco poles	Polycentric
Maclura tinctoria (L.) D. Don ex Steud.	Mac.tin	20	159	179	Colombian to Paranean poles	Polycentric

Nectandra megapotamica (Spreng.) Mez	Nec.meg	44	47	91	Paranean pole	Monocentric
Phyllostylon rhamnoides (J. Poiss.) Taub.	Phy. rha	0	28	28	São Francisco pole	Monocentric
Schinopsis balansae Engl.	Sch.bal	35	24	59	Chacoan pole	Monocentric
Schinopsis quebracho-colorado (Schltdl.) Barkley & Mey.	Sch.que	26	28	54	Chacoan pole	Monocentric
Acacia praecox Griseb.	Aca.pra	25	0	-	-	-
Astronium	Ast.fra	9	73	-	-	-

fraxinifolium Schott						
Balfourodendron riedelianum (Engl.) Engl.	Bal.rie	32	15	-	-	-
Bulnesia sarmientoi Griseb.	Bul.sar	9	0	-	-	-
Cabralea canjerana (Vell.) Mart.	Cab. can	0	48	-	-	-
Calycophyllum multiflorum Griseb.	Cal.mul	38	27	-	-	-
Caryocar brasiliense Cambess.	Car. bra	0	34	-	-	-

Capparis retusa Griseb.	Cap.ret	37	0	-	-	-
Capparis speciosa Griseb.	Cap.spe	13	0	-	-	-
Cercidium praecox (Ruiz & Pav.) Harms	Cer.pra	16	0	-	-	-
Chrysophyllum gonocarpum (Mart. & Eichler) Engl.	Chr.gon	17	54	-	-	-
Chrysophyllum marginatum (Hook. & Arn.) Radlk.	Chr.mar	36	60	-	-	-
Cochlospermum regium (Schrank)	Coc.reg	13	34	-	-	-

Pilg.						
Diplokeleba floribunda N. E. Br.	Dip.flo	37	11	-	-	-
Geoffroea decorticans (Hook. & Arn.) Burkart	Geo.dec	9	65	-	-	-
Jacaranda cuspidifolia Mart.	Jac.cus	11	38	-	-	-
Patagonula americana L.	Pat.ame	21	25	-	-	-
Peltophorum dubium (Spreng.) Taub.	Pel.dub	11	40	-	-	-

Pradosia brevipes (Pierre) T. D. Penn.	Pra. bre	0	14	-	-	-
Prosopis alba Griseb.	Pro.alb	11	48	-	-	-
Prosopis nigra (Griseb.) Hieron.	Pro.nig	14	33	-	-	-
Ruprechtia triflora Griseb.	Rup. tri	0	9	-	-	-
Sorocea bonplandii (Baill.) W.C. Burger, Lanj. & Wess. Boer	Sor.bon	27	39	-	-	-
Tabebuia heptaphylla (Vell.) Toledo	Tab.hep	18	38	-	-	-

Tabebuia nodosa (Griseb.) Griseb.	Tab.nod	19	59	-	-	-
Trichilia elegans A. Juss.	Tri.ele	21	49	-	-	-
Xylopia aromatica (Lam.) Mart.	Xyl. aro	0	18	-	-	-

Table 2

Theme	Min. value	Max. value	Unit	Reference
Mean precipitation per month	0	205.5	Mean/ grid (mm)	IPCC
Mean number of wet days per year	12.5	260.6	Mean/grid (nb.)	IPCC
Minimum temperature in July	-1	25	Mean/grid (C°)	IIASA Climate
Number of frost days in July	0	30	Number of frost days/grid (nb.)	IRI
Mean yearly temperature	-27	30	Mean/grid (C°)	IIASA Climate
Altitude	1	8000	Maximum/ grid (m)	GTOPO30

Figure Legends

Figure 1A Map of Paraguay; the main rivers and climatic areas are indicated. **B** Map of South America showing main biogeographic areas cited in the text.

Figure 2A Typology of the occurrences on the first factorial plane. Six vegetation communities (A0-2, B0-2) with a rather homogeneous floristic composition have been visually defined. Each group is identified by a star connecting all the occurrences of the group to its barycentre on the factorial plane. **B** Geographical position of the six vegetation communities defined by DAENO analysis. Each community is identified by a star connecting the occurrences to the geographical barycentre of the group. The contour polygon of each group is also displayed (Light grey polygons correspond to the chacoan biome, and dark grey polygons indicate the Paranean biome). **C** Species composition of the six groups of occurrences defined by DAENO analysis (See Table 1 for species abbreviations). The importance of a species in a given community is represented by a black square. The square size is proportional to the percentage of the total number of occurrences of the community that is represented by this species.

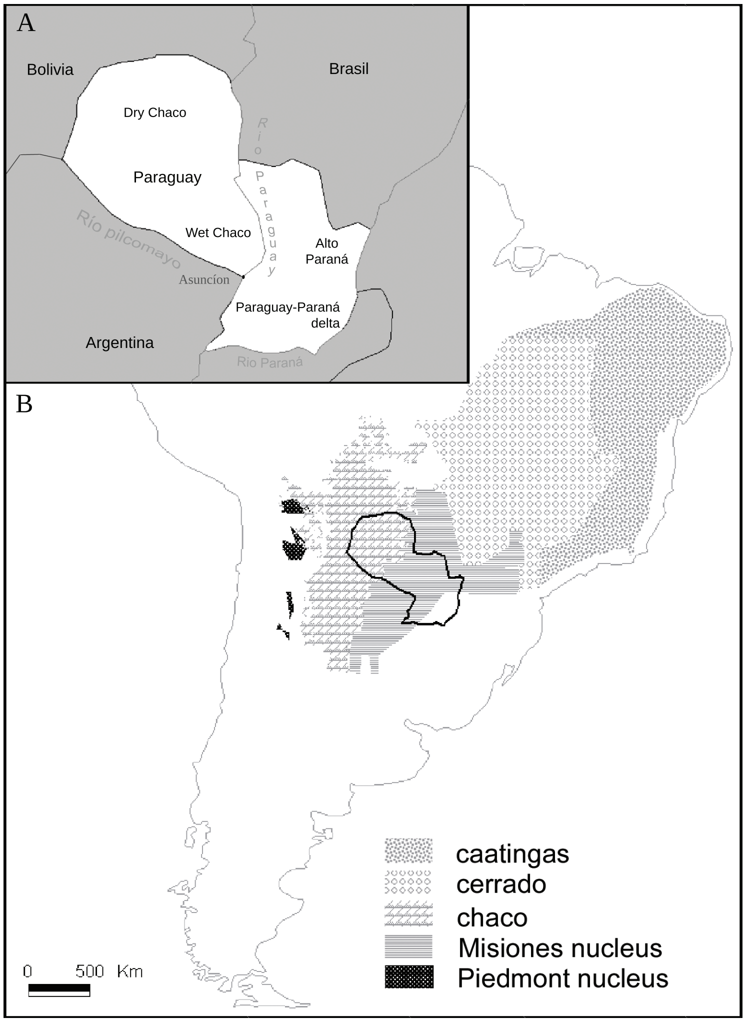
Figure 3 Factorial map of the occurrence scores on the second and third axes of the discriminant analysis of the eigenvectors of the Neighbourhood Operator. Categorization of the tree occurrences in four classes according to their position in the three dimensional space defined by the first three axes: Chacoan, Colombian, São Francisco and Paraná poles (Spichiger et al., 2004, Figure 3b, Blackwell Publishing).

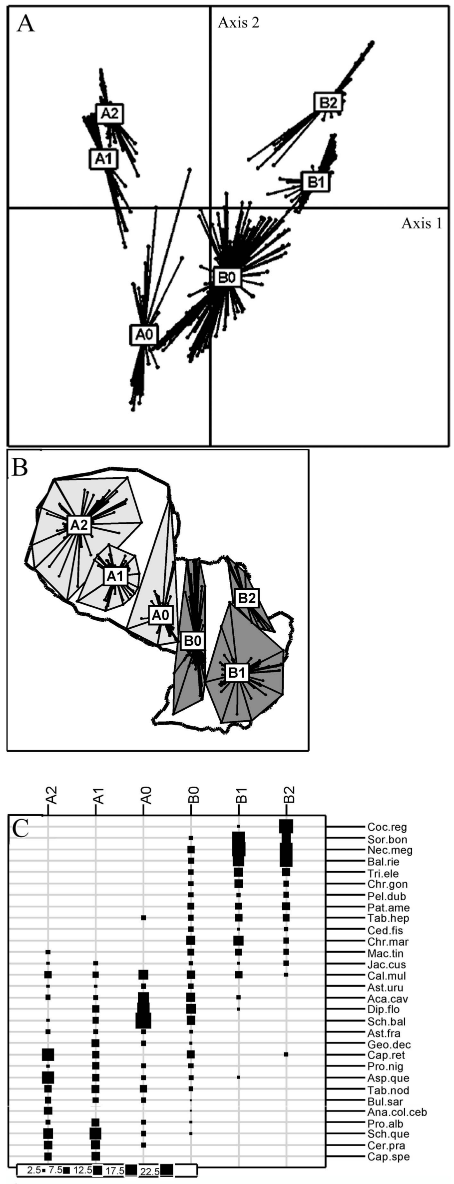
Figure 4 Potential and present-day distributions of two Paranean and two Chacoan species in an equilibrium situation.

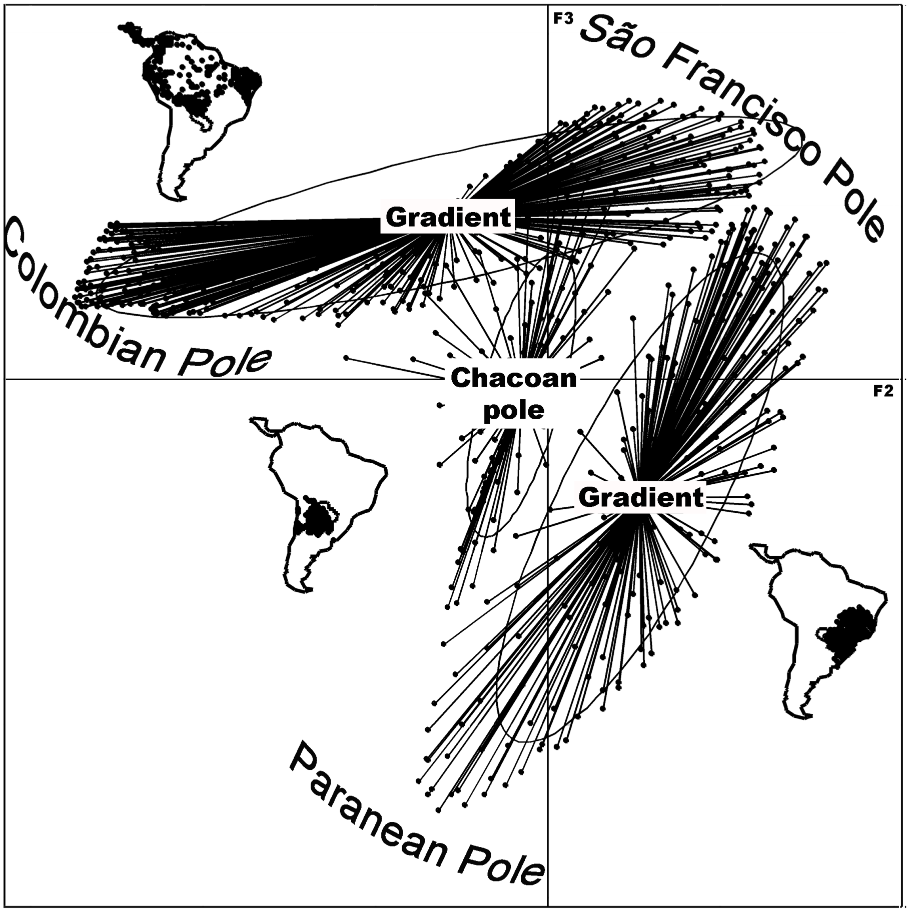
Figure 5 Potential and present-day distributions of four polycentric species in a disequilibrium situation.

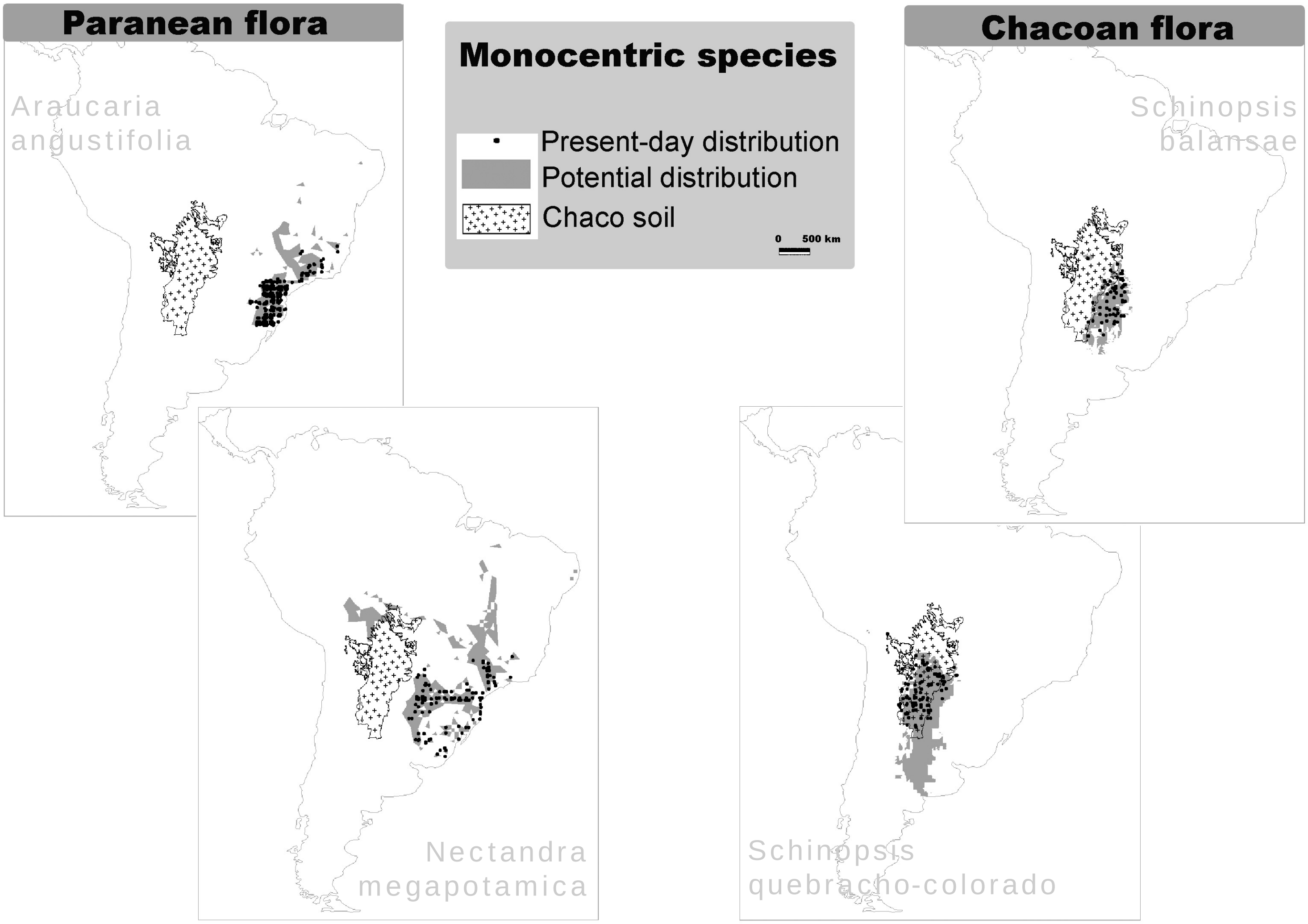
Figure 6 Past (LGM) potential distributions of polycentric, Paranean and chacoan species, illustrated with Patagonian steppic and montane elements (Markgraf, 1991; Pennington et al., 2000). The Amazon basin is invaded by Paranean and polycentric species and the Paraná basin by steppic and montane elements.

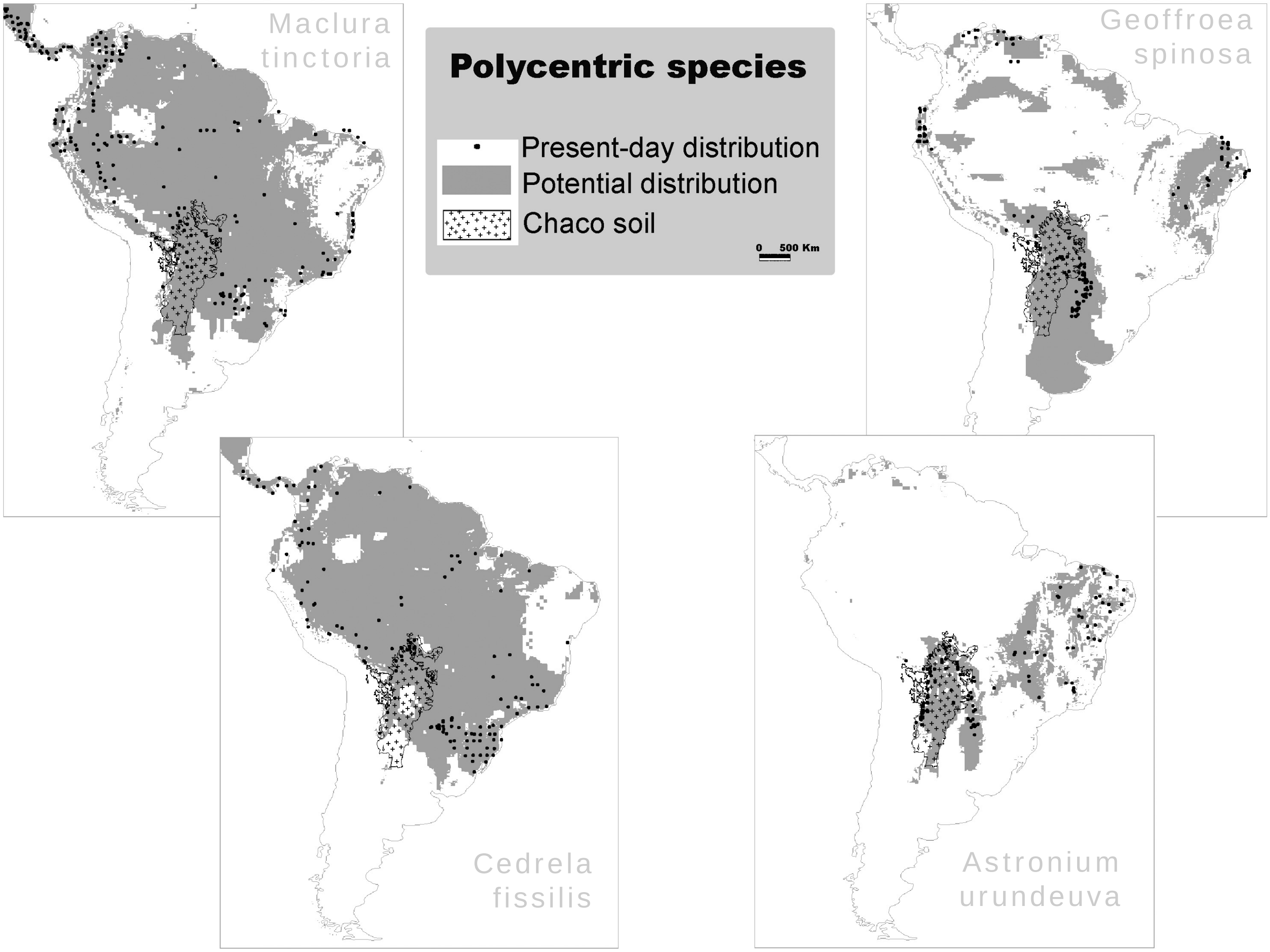
Figure 7 Main ecological trends affecting the present Paraguayan tree-species. **A:** alkaline soil, loamy texture, low rainfall, corresponding to "chaco seco" flora. **B:** acid soil, sandy texture, low rainfall, corresponding RPSD flora. **C:** acid soil, clayey texture, high rainfall, corresponding to Paranean flora (Spichiger et al., 1995).

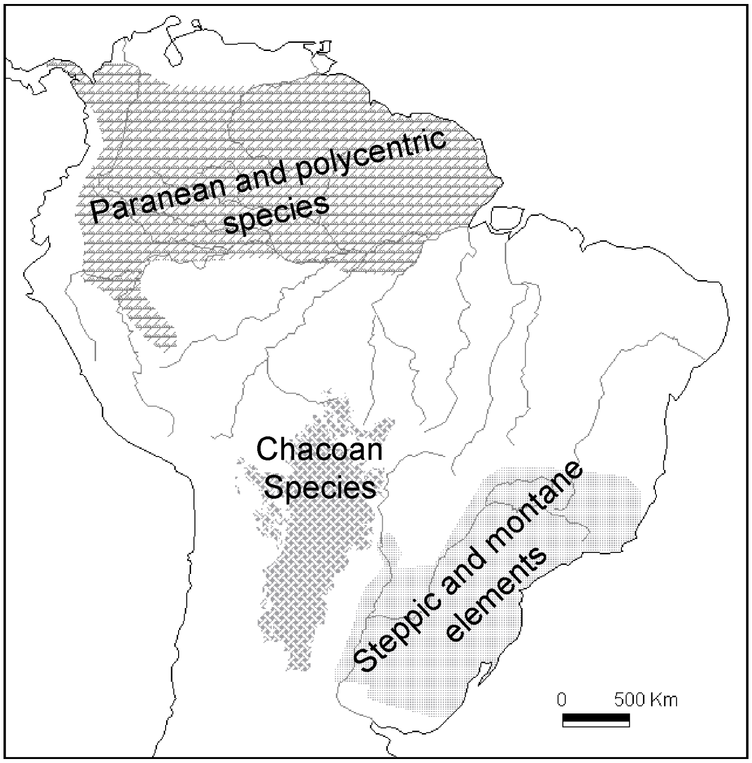

A
Bolivia
Brasil
Dry Chaco
Paraguay
Río Paraguay
Wet Chaco
Alto Paraná
Río pilcomayo
Asunción
Paraguay-Paraná delta
Argentina
Río Paraná
B
caatingas
cerrado
chaco
Misiones nucleus
Piedmont nucleus
0 500 Km

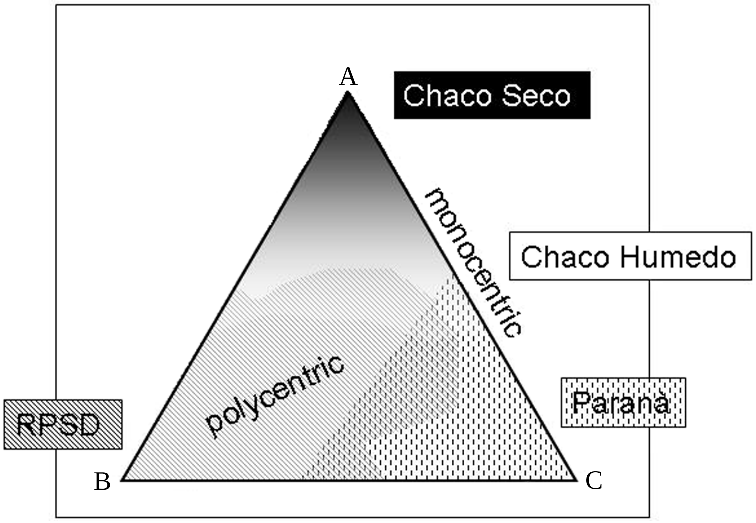

A
Axis 2
Axis 1
A2
A1
A0
B0
B1
B2
B
A2
A1
A0
B0
B2
B1
C
A2
A1
A0
B0
B1
B2
Coc.reg
Sor.bon
Nec.meg
Bal.rie
Tri.ele
Chr.gon
Pel.dub
Pat.ame
Tab.hep
Ced.fis
Chr.mar
Mac.tin
Jac.cus
Cal.mul
Ast.uru
Aca.cav
Dip.flo
Sch.bal
Ast.fra
Geo.dec
Cap.ret
Pro.nig
Asp.que
Tab.nod
Bul.sar
Ana.col.ceb
Pro.alb
Sch.que
Cer.pra
Cap.spe
2.5
7.5
12.5
17.5
22.5


F3
São Francisco Pole
Gradient
Colombian Pole
Chacoan pole
F2
Gradient
Paranean Pole


Paranean flora
Monocentric species
Chacoan flora
Araucaria angustifolia
Present-day distribution
Potential distribution
Chaco soil
0 500 km
Schinopsis balansae
Nectandra megapotamica
Schinopsis quebracho-colorado


Polycentric species
Present-day distribution
Potential distribution
Chaco soil
0 500 Km
Maclura tinctoria
Geoffroea spinosa
Cedrela fissilis
Astronium urundeuva


Paranean and polycentric species
Chacoan Species
Steppic and montane elements
0
500 Km


A
Chaco Seco
monocentric
Chaco Humedo
polycentric
Paraná
RPSD
B
C

Annexe 7

Article 7 :

“The Kernel method for the estimation of the geographical distribution of species occurrences”

Clément Calenge, Cyrille Chatelain & Rodolphe Spichiger.

Applied Vegetation Science, soumis

The Kernel method for the estimation of the geographical distribution of species occurrences

Calenge, Clément[1*]; Chatelain, Cyrille[2] & Spichiger, Rodolphe[3]

[1*] *Laboratoire de Biométrie et Biologie Evolutive, Université Claude Bernard Lyon 1, 43 Boulevard du 11 novembre 1918, 69622 Villeurbanne Cedex, France*; [2] *Conservatoire et Jardin botaniques de la Ville de Genève, Ch. de l'Impératrice 1, CP 60, CH-1292 Chambésy, Genève, Switzerland; E-mail Cyrille.Chatelain@cjb.ville-ge.ch;* [3] *Conservatoire et Jardin botaniques de la Ville de Genève ; E-mail Rodolphe.Spichiger@cjb.ville-ge.ch*

[*] *Corresponding author; Fax +33 0472431388 ; E-mail calenge@biomserv.univ-lyon1.fr*

Abstract. Herbarium data contain a lot of information on the spatial distribution of species, and for this reason, herbarium specimens are a very common data source in conservation biology and biogeography studies. When a specimen – a species occurrence – is collected, information on its location is also noted. In this paper, we describe the mathematical principle of the kernel estimator, a smoothing method that can be used to estimate the "Geographical Distribution" (GD) of a species from the distribution pattern of its occurrences. We define here the GD of a species as a bivariate probability density function giving the density of probability of a species occurrence at a given place, according to its geographical coordinates. This nonparametric method is quite accurate, relying only on one smoothing parameter. This estimator can be used to discriminate between monocentric species (species with a non-fragmented distribution) and polycentric species (species with fragmented distribution), or more simply to explore the shape of the GD. In addition, once this estimator has been fitted to the data, several numerical parameters describing the shape or the size of the species distribution can be derived (e.g. the "distribution range"). We stress however that the kernel method implies the crucial hypothesis of a uniform sampling effort over the studied area. We present an example of application of the method using two tree-species *Cedrela fissilis* Vell. and *Schinopsis quebracho-colorado* Schltdl. growing in South America

Keywords: Kernel estimator; spatial distribution; herbarium data; mapping; point pattern, R package; South America.

Nomenclature: Spichiger & Ramella (1998)

Abbreviations:

AOO = Area Of Occurrence
DR = Distribution Range
EOO = Extent Of Occurrence
GD = Geographical Distribution
LSCV = Least Square Cross Validation
MISE = Mean Integrated Square Error
pdf = probability density function

Introduction

The description of the geographic distribution of a species is a central issue in biogeography. It is used to determine the endangered status of the species (Gaston 1994; IUCN 1994), or to gain knowledge on its paleoecology (Spichiger et al. 2004). The data used to analyse this distribution generally consists of records of the taxon at specified locations, which are usually termed "occurrences" in the literature (Gimaret et al. 2003). The occurrence locations are usually derived from herbarium specimens and their associated locality information. In most studies, it is supposed that the occurrences distribution reflects the actual distribution of the species (Willis et al. 2003).

In some studies (e.g. taxonomic reviews), the species distribution is described only with the help of the "crude" occurrence patterns. Other studies derive parameters measuring features of this distribution (Gautier et al. 1999; Raxworthy et al. 2003). In conservation studies two main measures, the Extent Of Occurrence (EOO) and the Area Of Occurrence (AOO) (IUCN 1994), are generally used to determine the status of endangered species,. More rarely, shape parameters are used to describe distribution. Thus, several authors recommend the measurement of the fragmentation, to describe the way in which habitats or subpopulations within a landscape are divided into large or small patches (Hartley & Kunin 2003; Willis et al. 2003). Whether size or shape of the distribution is considered, the parameters are often derived from the occurrence pattern in an *ad hoc* way. For example, the EOO is estimated by drawing the minimum convex polygon encompassing all occurrences. However, the occurrence distributions may not be convex, or may be fragmented into several subpopulations (polycentric species, see Willis et al. 2003; Spichiger et al. 2004), and the estimated EOO may contain large areas from which the species is absent. To cope with this problem, the Area Of Occupancy (AOO) of a species has been defined as the area actually occupied by the taxon, excluding cases of vagrancy (IUCN 1994). A virtual grid is superinposed on to the occurrences pattern, and the AOO corresponds to the sum of the areas of the grid cells containing at least one occurrence. However, the estimated area is strongly dependent on the size of the grid and on the number of occurrences at hand. In addition, both the AOO and the EOO do not reveal any information about the spatial pattern of distribution.

We introduce here a probabilistic definition of the species distribution, which renders both the exploration of the features of the occurrence pattern (for taxonomic reviews) and the estimation of parameters descriptive of this distribution easier. The set of species occurrences can be considered to be the realization of a bivariate probability density function (pdf). This function, which we call “Geographical Distribution” (GD), gives the probability density of collecting an occurrence at a given point according to its geographical coordinates (Fig. 1; see Worton 1989). The main focus here is to estimate this function with a random sample of occurrences, to allow a better description of the range of a species. Once this distribution has been estimated, several parameters describing the size and shape of the distribution can be derived in a straightforward way. We describe a smoothing method, the kernel estimator, which provide this estimation. This method is non-parametric (no parametric form is assumed for the GD) and can therefore be performed on any type of data, including very fragmented distributions. We provide an example of the application of the method on two south-American tree species.

The kernel method for the estimation of the density

The estimation of a pdf from a set of observations is generally carried out with the help of smoothing methods such as the kernel method. Here, we develop the principle of this estimator. To promote a better understanding of the method, we first consider the case of the estimation of the pdf of one single variable, and we then extend it to the estimation of the bivariate GD.

Univariate estimation of density

By definition, the probability density of a continuous random variable X is

$$f(x) = \lim_{h \to 0} \frac{1}{2h} P(x-h < X < x+h)$$

Let $X_1, \dots X_i, \dots X_n$ be a set of n independent observations of a continuous variable X. For any given h, we can estimate $\hat{P}(x-h < X < x+h)$ by the proportion of observations falling in the interval $[x-h, x+h]$. Then, the pdf of X can be estimated with

$$\hat{f}(x) = \frac{1}{2h} \hat{P}(x-h < X < x+h)$$

This "naïve estimator" (Silverman 1986) can be reformulated

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} w\left(\frac{x - X_i}{h}\right) \qquad \text{(Eq. 1)}$$

where

$$w(x) = \begin{cases} 1/2 & \text{if } |x|<1 \\ 0 & \text{otherwise} \end{cases}$$

The naïve estimator is therefore equivalent to an attempt to place a "box" of height $1/(2nh)$ and width $2h$ on each observation and then summing these boxes at a given point to obtain the estimate of the pdf at this point (Fig. 2a). However, this is a poor estimation because of its ragged aspect.

The kernel method is an extension of this naïve estimator that gives smoother estimates. The pdf is estimated by placing a "bump" of area 1/n, instead of a box, on each observation and then summing to obtain the estimate. In other words, the function w is replaced in Eq. 1 by a kernel function K which satisfies

$$\int_{-\infty}^{+\infty} K(x)\mathrm{d}x = 1$$

Usually, this function is a symmetric probability density function (such as the Gaussian distribution). The estimation does respect all properties of a pdf, i.e. it is continuous, always positive and integrates to 1 (Fig. 2b).

Bivariate case

The above mathematical bases may be extended to fit a pdf in a higher dimensional space (Silverman 1986). We consider here a sample of n occurrences distributed over a non bounded two-dimensional space, and we want to estimate the probability density of collecting an occurrence at a given point according to its geographical coordinates X and Y. We therefore develop here the extension of the kernel method to the two-dimensional space.

Let $\mathbf{U}_1$, … $\mathbf{U}_\mathrm{i}$, … $\mathbf{U}_\mathrm{n}$ be a set of n independent occurrences of a given species. The vector $\mathbf{U}_\mathrm{i} = (X_i, Y_i)$ contains the coordinates of the occurrence i. The kernel estimation of the GD at point $\mathbf{u} = (x, y)$ is computed using the following equation

$$\hat{f}(\mathbf{u}) = \frac{1}{nh^2}\sum_{i=1}^{n} K\left\{\frac{1}{h}(\mathbf{u}-\mathbf{U}_i)\right\} \text{ (Eq. 2)}$$

where K is a unimodal symmetrical bivariate probability density function, and h is the smoothing parameter that can be varied by the user. The choice of a kernel function is not crucial, as all possible kernel functions have a close efficiency, i.e. similar bias and variance. One common kernel function is the bivariate normal density. In this case, the kernel estimator of the geographical distribution is

$$K(\mathbf{u}) = \frac{1}{nh^2} \sum_{i=1}^{n} \frac{1}{2\pi} \cdot \exp\left(- \frac{(\mathbf{u} - \mathbf{U}_i)^{\mathrm{T}} (\mathbf{u} - \mathbf{U}_i)}{2h^2} \right)$$

, where $\mathbf{x}^{\mathrm{T}}$ is the transpose of $\mathbf{x}$. The estimation of the GD at any point $\mathbf{u}$ is therefore a function of the distance between this point and the occurrences. The closer the occurrences are, and the higher the value of the GD at this point is. Again, this is equivalent to an attempt to place a three-dimensional "bump" of volume $1/n$ on each observation and then summing these to obtain the estimate. Practically, the GD is estimated at each node of a virtual grid superimposed on the study area. The size of the grid does not really affect the estimate (Worton 1995).

Choosing a value for h

The most sensible parameter for the estimation is the smoothing parameter h, also called bandwidth or window width. The smoothing parameter controls the width of the "bump" placed on each observation, and therefore has a major influence on the appearance of the fitted GD. However, as noted by Silverman (1986), "it should never be forgotten that the appropriate choice of smoothing parameter will always be influenced by the purpose for which the density estimate is to be used. If the purpose of density estimation is to explore the data in order to suggest possible models and hypotheses, then it will probably be quite sufficient, and indeed desirable, to choose the smoothing parameter subjectively".

However, as noted above, precise parameters are sometimes derived from the estimated GD (e.g. distribution range, number of core areas, shape parameters, see below), and the estimation is required to be precise. One criterion measuring the goodness of fit of the estimation is the Mean Integrated Square Error (MISE), which measures the error of the estimation between the estimated and the true pdf, for a given value of *h*:

$$MISE(h) = E\int(\hat{f}_h - f)^2$$

where $\hat{f}_h$ is the pdf estimated with the smoothing parameter *h* and *f* is the "true" GD. The best value of *h* is the value that minimizes *MISE*(*h*). However, the true distribution is often unknown, and the MISE cannot be computed. Two main solutions may be proposed.

One possibility to cope with this problem is to consider that the true density belongs to a family of standard distribution. A popular approach is to assume that the "true" geographical distribution is bivariate normal, with the variance covariance matrix (V):

$$\mathbf{V} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}.$$

Theoretical developments (Silverman 1986) lead to the estimation

$$\hat{h}_{opt} = \hat{\sigma} n^{-1/6}$$

where $\hat{\sigma} = \left\{ \frac{1}{2}\left[\hat{\sigma}_X^2 + \hat{\sigma}_Y^2\right]\right\}^{\frac{1}{2}}$, and $\hat{\sigma}_X^2$ and $\hat{\sigma}_Y^2$ are the estimated variances for the coordinates *X* and *Y* of the occurrences. This *ad hoc* choice for *h* is widely used and often gives good results (Andreassen et al. 1993). However, when the GD is suspected to be multimodal (i.e. several subpopulations separated by vast areas), the *ad hoc* choice for h leads to over-smoothed estimates (Seaman et al. 1999).

Another possibility is to choose h by Least Square Cross-Validation (LSCV). This estimation minimizes the discrepancy between the estimate and true GD. The value of h that minimises the MISE also minimises the function

$$CV(h) = \frac{1}{\pi h^2 n} + \frac{1}{4\pi h^2 n^2} \times \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\exp\left[-\frac{d_{ij}^2}{4h^2} \right] - 4\exp\left[-\frac{d_{ij}^2}{2h^2} \right] \right)$$

where d_{ij} is the distance between the occurrence i and the occurrence j. The minimum value of h is calculated by plotting $CV(h)$ as a function of h to determine the minimum. The value of h that minimises $CV(h)$ is denoted h_{CV}. This value leads to estimated GD that closely fit the data. On the other hand, this method leads to strongly over-smoothed GD when the number of occurrences is small (<30, see below).

Deriving the distribution range from the geographical distribution

The kernel method is often used to derive a large number of parameters descriptive of the size or the shape of the GD. One useful tool is a measure of the area of occupancy, which we term “Distribution Range” (DR), by analogy with the “home-range” of animals monitored using radio-tracking (Worton 1989). The DR is the minimum area in which a species occurrence has some specified probability of being collected. Thus, the DR at 95% of a given species corresponds to the smaller area where an occurrence of this species has a probability of 0.95 of being located. Alternatively, DRs estimated for lower probabilities, e.g. 50%, identify areas where the species is the most frequent. The estimated limits of such “core areas” are more precise than the boundaries of DR estimated at larger levels, because the kernel estimator leads to estimation with highly variables tails (Anderson 1982; Silverman 1986).

Although the visual examination of the GD may bring very useful information, numerous quantifiable parameters can be derived from the DR. Thus, the DR size may give an index of the size of the GD. As the smoothing parameter can be determined in an objective way (e.g. using the LSCV method), the DR size at a given level (e.g. 50%) can be used in conservation studies to give an index of the size of the distribution, in addition to AOO and EOO. If the GD is fragmented, the number of parts of the DR indicates whether the species is monocentric or polycentric. Numerous other parameters can be derived from the estimated DR (for examples see Ford 1983; Andreassen et al. 1993).

Practical use of the method

We describe here two applications of the kernel method in the fields of biogeography and conservation biology. We use South American tree species to illustrate the method. We first use the kernel estimator to derive the GD of *Cedrela fissilis* Vell., and to draw hypotheses about the past distribution of the species. We also consider the possible use of the GD to explore species conservation opportunities. Then, we estimate the distribution limits of *Schinopsis quebracho-colorado* Schltdl., a tree species growing only on the salty soils of the Chaco. Because this species has a well-defined distribution in South America, the estimated distribution of the species has been compared with the actual distribution, to point out the limitations of the method. For each species, a review of the literature (monographs and taxonomic reviews) allowed us to ascertain the presence of the focus species at many locations on the continental scale (121 locations for *C. fissilis*, and 48 locations for *S. quebraco-colorado*). These presences have been located on a map with the help of a Geographic Information System (ESRI 1996). These data have then been exported into R software (Ihaka & Gentleman 1996), where the kernel estimator has been fitted to the data.

The distribution of Cedrela fissilis: *biogeography and conservation issue*

Mueller (1973) noted that "Every species possesses, or used to possess, at least one dispersal centre that was its centre of origin. During the evolution of the taxon, however, the centre of origin and the centre of dispersal can become widely separated from each other." The identification for a given species of the "centres", or nuclei, of the present distribution, is essential in biogeography. The comparison of the environmental characteristics (temperature, type of soil, etc.) of these nuclei and the rest of the study area may reveal the essential characteristics for the growth of the species and its ecological amplitude. In addition, the nuclei of a fragmented distribution have necessarily been connected in the past, and the space between the separated centres was probably once colonised by the species.

A recent work (Spichiger et al. 2004) has shown that Paraguayan tree species may be found either in only one (monocentric species) or in several centres (polycentric species). The kernel method allows us to enhance the location of the nuclei of the distribution of *C. fissilis* and to determine whether this species is monocentric or polycentric, and thereby develop hypotheses about the past distribution of this species.

The *ad hoc* choice for the smoothing parameter $h_{\text{opt}} = 6.93°$, leads to a strongly over-smoothed estimation (Fig. 3b). Indeed, the distribution appears to be multimodal (several centres), which violates the assumption of unimodality underlying the *ad hoc* choice of *h*. However, this estimation is good enough to show the main characteristics of the species distribution. The 50% DR is made of two nuclei. The main subpopulation is located in the Paraná (Paraguay-Uruguay) and the second one is located in the north of the continent (Colombia-Venezuela-Peru).

Because of the strong over-estimation of the distribution, we performed the analyses again with the value of the smoothing parameter minimising the LSCV criterion ($h_{\text{CV}} = 1.93°$; Fig. 3c). This value, smaller than the *ad hoc* choice of *h*, results in an estimation that closely fits the data (Fig. 3d). In addition, the fine structures of the distribution are revealed by this estimation. A third nucleus resolves with this value of *h*. This nucleus is located at the mouth of the Amazon River.

Cedrela fissilis is a generalist species (Spichiger et al. 2004), i.e. it can grow under a wide variety of conditions. The huge tracks of land between the three nuclei indicate that this species has probably been present on the whole continent in the recent past. This would confirm the hypothesis of a Holocene (~10,000 years B.P.) rise in both temperature and rainfall, postulated by many authors (Prado & Gibbs 1993; Pennington 2000; Behling 2002). Climate variation may have caused the drought-tolerant and generalist species such as *C. fissilis* to move to the edges of the Amazon Basin and be replaced in the wettest areas by drought-sensitive elements (for further details, see Spichiger et al. 2004).

The kernel method may also be useful in Conservation biology. Although *C. fissilis* is rather widespread, it is listed as endangered on the IUCN red list. Understanding biogeographical and ecological characteristics of plants species is very important to plan conservation strategies. When the GD is fragmented into several nuclei, their location may help to identify the causes of fragmentation of the distribution, and therefore to plan a conservation policy to act on these causes. On the other hand, knowing the location of these nuclei may help to indicate the areas where the species are to be preserved (e.g. by the creation of reserves). For example, the south-eastern nucleus of the GD of *C. fissilis* is made of two clearly separated "sub-nuclei," one large sub-nucleus centred on the Paraná region (South-Paraguay, Uruguay), and one smaller centred in the south of Bolivia (see Fig. 3d). The area without occurrences in the Chaco region (north-west of Paraguay), between these two areas, is evident. These two populations are probably in contact through the net of riverine forests found in the Chaco area (Spichiger et al. 2004). This would stress the need to preserve these forests, to allow the preservation of the south Bolivian nucleus of *C. fissilis* by allowing genetic exchange with the larger Paranean sub-nucleus. Other fragmented nuclei are found for the distribution of this species. The nucleus of the north of South America seems clustered into one main sub-nucleus in Colombia-Peru, and one smaller sub-nucleus in Venezuela. Visual or automatic identifications of such disjunctions are problematic with the classical plant AOO approach, as it depends mainly on the size and position of the grid used (Fig. 4). Using the kernel method, we could consider how one could pay attention to each small isolated nucleus in a conservation strategy.

Limits of the method: the distribution of Schinopsis quebracho-colorado

We focus here on *Schinopsis quebracho-colorado*, a species for which the true distribution is known. This species only grows on the salty soils of the Chaco, a dry region in South America. The Chaco extends from northern Argentina to southern Bolivia, and through western Paraguay. This area is characterized by loamy-sandy xeric soils, the scarcity of precipitation and by the occurrence of frost, and has well-defined boundaries (Fig. 5a). The estimated distribution of *S. quebracho-colorado* should be totally included within the Chaco boundaries. We used the kernel method to estimate the GD of this species, and we then measured the percentage of the 95% DR and of the 50% DR that is included within the Chaco region.

The choice of the smoothing parameter was not crucial in this case, as both the *ad hoc* choice and the LSCV led to similar values ($h_{\mathrm{opt}} = 1.05°$ vs. $h_{\mathrm{CV}} = 0.805°$). We chose the *ad hoc* value, keeping in mind that both choices are equivalent. Only 59% of the 95% DR was included inside the Chaco (Fig. 5a). In this case, a larger part of the minimum polygon encompassing all occurrences was included within the boundaries of the Chaco (70.1%; Fig. 5b). Thus, a large part of the estimated distribution was located outside the Chaco. Even the 50% DR was not totally included within the Chaco boundaries (96.4%).

A careful examination of the distribution may explain why this estimation was so poor. First, the occurrence locations were rounded to the nearest degree of both latitude and longitude, so that we did not have access to the true locations of the occurrences. Several occurrences were located outside the boundaries of the Chaco (20% of the occurrences). The number of occurrences was small, so that the estimated GD was strongly over-smoothed, even with the *ad hoc* choice for *h*. Finally, this lack of fit may be explained by a “boundary effect.” By definition, the estimated distribution is a probability density function that decreases asymptotically to zero as the distance from the centre of the distribution increases. Therefore the model for the GD does not take into account the possibility of a physical or ecological boundary to the distribution. Only the core of the distribution (e.g. DR at 50%) is precisely estimated.

Discussion

We presented here a method to estimate the geographic distribution of a species with occurrence data. This technique relies on a probabilistic definition of the species distribution. Because it is nonparametric, this method is very robust and can be applied to any kind of distribution, even on very fragmented distributions. However, we have shown that the method does not take into account the physical boundaries present in the study area, since it supposes an unbounded space. Other smoothing methods are available to fit the GD to a set of occurrences (e.g. lowess regression; Cleveland & Devlin 1988), however, we recommend the use of the kernel method because of its widely known properties. The kernel estimator is the more commonly used technique, and it is probably the method whose properties are best understood (Silverman 1986). This method is frequently used in other fields of Ecology, and especially Wildlife Ecology. The kernel method has long been used to estimate the home ranges of animals monitored by radio-tracking (Worton 1989). Numerous empirical and simulation studies have been carried out to determine the practical efficiency and the limitations of the method (e.g. Worton 1995; Seaman et al. 1999).

The sample size required by the kernel method depends on the objectives of the study. As we saw above, low sample size does not allow precise estimates of the limits of a distribution. Simulation studies have shown that a low number of occurrences may lead to strongly over-estimated DR size (over-smoothing of the GD; Seaman et al. 1999). In this case, a minimum sample size of 30 occurrences is generally recommended to achieve this objective, and preferably more than 50. However, if the objective is to explore the main features of the occurrence distribution, lower sample sizes are in our opinion acceptable.

Another issue frequently reported in the literature is the sensitivity of the kernel estimator to outliers (cases of vagrancy; Silverman 1986). When outliers occur in the data, the estimated GD tends to be over-smoothed in areas where the species occurs in high density, whereas in low density areas, it tends to be under-smoothed. The kernel method may therefore obscure useful details of the core of the GD and reveal random noise in the tail of the GD. One possibility may be to exclude cases of vagrancy before performing the analysis. Another possibility is to use the adaptive kernel method, a method developed to cope with this problem (Silverman 1986). With this technique, the smoothing parameter h varies according to the local density of occurrences, being smaller in regions of high density and larger in regions of weak density. This method is rarely used, as several authors indicated that it returns biased results (Worton 1989; Seaman et al. 1999). However, we stress that further studies are needed to explore more deeply the practical properties of this promising method for application in biogeography.

The most crucial issue concerning the kernel estimator does not relate to the method, but rather to the data. In biogeography, the data on the distribution generally originates from herbarium samples. Thus, what we termed “Geographical Distribution” estimates the probability density of collecting an occurrence, which is not the same as the probability density of presence of the species. The GD is a function of both the actual species distribution and the distribution of the sampling effort. When the kernel method is to be used for inference about the actual distribution of the species, the crucial hypothesis is that the sampling effort is the same all over the area. The kernel method shares this property with all other methods used for estimating the species distribution or area of occupancy. We refer the reader to Willis et al (2003) for a deeper discussion on this issue.

We think that Kernel estimator may provide information useful to biologists and biogeographers and bring a method less dependant on scale than the AOO or EOO. It formalizes several general concepts within a probabilistic framework. We described the distribution range, the core areas and the shape parameter *S*, but many other parameters describing size and shape of the distribution can be derived from the GD, according to the objective of the studies. More than any other method, the kernel estimator may serve to explore the internal configuration of a geographical distribution, and help biologists who want to develop hypotheses on the history, ecology or conservation of the species.

Software availability

Numerous functions in the R software (Ihaka & Gentleman 1996) are available to perform kernel estimation. The package *adehabitat* contains several functions to perform this estimation, and especially the estimation of distribution ranges. Both the package and the R software can be downloaded freely at the URL http://cran.r-project.org/.

Literature Cited

Anderson, D.J. 1982. The home range: a new nonparametric estimation technique. *Ecology* 63: 103-112.

Andreassen, H.P., Ims, R.A., Stenseth, N.C. & Yoccoz, N.G. 1993. Investigating space use by means of radiotelemetry and other methods: a methodological guide. In: Stenseth, NC & Ims, RA (eds.) *The biology of Lemmings*, pp. 589-618. Academic press, London, UK.

Behling, H. 1993. Untersuchungen zur spätpleistozänen und holozänen Vegetations- und Klimageschichte der tropischen Küstenwälder und der Araukarienwälder in Santa Catarina (Südbrasilien). *Diss. Bot.* 206: 1-149.

Cleveland, W.S. & Devlin, S.J. 1988. Locally weighted regression: an approach to regression analysis by local fitting. *J. Amer. Statist. Assn.* 83: 596-610.

Hartley, S. & Kunin, W.E. 2003. Scale dependency of rarity, extinction risk, and conservation priority. *Conserv. Biol.* 17: 1559-1570.

ESRI 1996. *Using Arcview GIS. The geographic information system for everyone*. Environmental Systems Research Institute, Inc, USA.

Ford, R.G. 1983. Home range in a patchy environment: optimal foraging prediction. *Am. Zool.*, 23: 315-326.

Gaston, K.J. 1994. *Rarity*. Chapman & Hall, London, UK.

Gautier, L., Aké Assi, L., Chatelain, C. & Spichiger, R. 1999. Ivoire: a geographic information system for biodiversity management in Ivory Coast. In: Timberlake, J. & Kativu, S. (eds.) *African plants: Biodiversity, taxonomy and uses*. pp. 183-194. Royal Botanic Gardens, Kew, UK.

Gimaret-Carpentier, C., Dray, S. & Pascal, J.P. 2003. Broad scale biodiversity pattern analyses of the endemic tree flora of the western Ghats (India) using canonical correlation analysis of herbarium records. *Ecography* 26: 429-444.

Ihaka, R. & Gentleman, R. 1996. R: A Language for Data Analysis and Graphics. *J. Comput. Graph. Stat.* 5: 299-314.

Mueller, P. 1973. *The dispersal centres of terrestrial vertebrates in the Neotropical realm: A study in the evolution of the Neotropical biota and its native landscapes*. Junk, La Hague, France.

Pennington, T.D. 2000. Neotropical seasonally dry forests and Quaternary vegetation changes. *J. Biogeogr.* 27: 261-273.

Prado, D.E. & Gibbs, P.E. 1993. Patterns of species distributions in the dry seasonal forests of South America*.* *Ann. Missouri Bot. Gard.* 80: 902-927.

Raxworthy, C.J., Martinez-Meyer, E., Horning, N. , Nussbaum, R.A., Schneider, G.E., Ortega-Huerta, M.A., & Peterson, A.T. 2003. Predicting distributions of known and unknown reptile species in Madagascar. *Nature* 426: 837-841.

Seaman, D.E., Millspaugh, J.J., Kernohan, B.J., Brundige, G.C., Raedeke, K.J. & Gitzen, R.A. 1999. Effects of sample size on kernel home range estimates. *J. Wild. Manage.* 63: 739-747.

Silverman, B.W. 1986. *Density estimation for statistics and data analysis.* Chapman & Hall, London, UK.

Spichiger, R. & Ramella, L. 1998. *Flora del Paraguay*. Edition des Conservatoires et Jardins botaniques de la Ville de Genève, Missouri Botanical Garden, Geneva, Swiss.

Spichiger, R., Calenge, C. & Bise, B. 2004. The geographical zonation in the Neotropics of tree species characteristic of the Paraguay-Paraná Basin, *J. Biogeogr. in press.*

UICN. 1994. IUCN Red List Categories. Prepared by the IUCN Species Survival Commission. IUCN, Gland.

Willis, F., Moat, J. & Paton, A. 2003. Defining a role for herbarium data in Red List assessments: a case study of Plectranthus from eastern and southern tropical. *Biodivers. Conserv.* 12: 1537-1552.

Worton, B.J. 1989. Kernel methods for estimating the utilization distribution in home-range studies. *Ecology* 70: 164-168.

Worton, B.J. 1995. Using Monte Carlo simulation to evaluate kernel-based home range estimators. *J. Wild. Manage.* 59: 794-800.

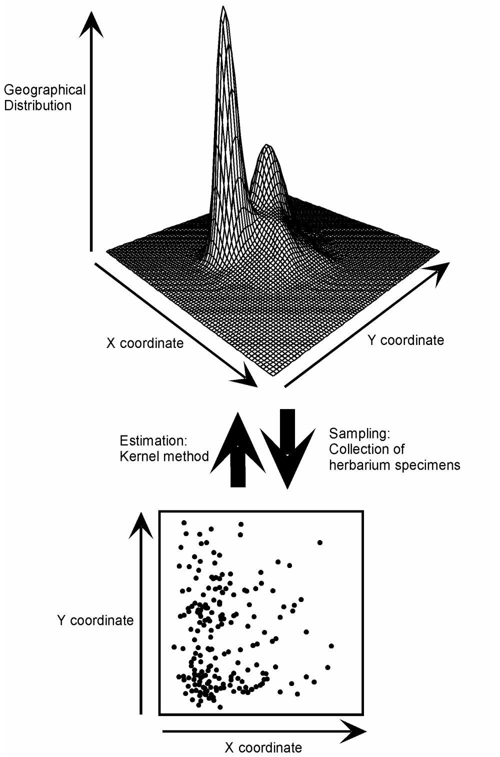


Fig. 1. The Geographical Distribution (GD) is considered to be a bivariate function giving the probability density of collecting an occurrence at a given place, according to its geographical coordinates. The realization of this function (sampling) gives a set of occurrences, e.g. through the collection of herbarium specimens. The main issue addressed by biogeographers is the estimation of the GD from the occurrence distribution.

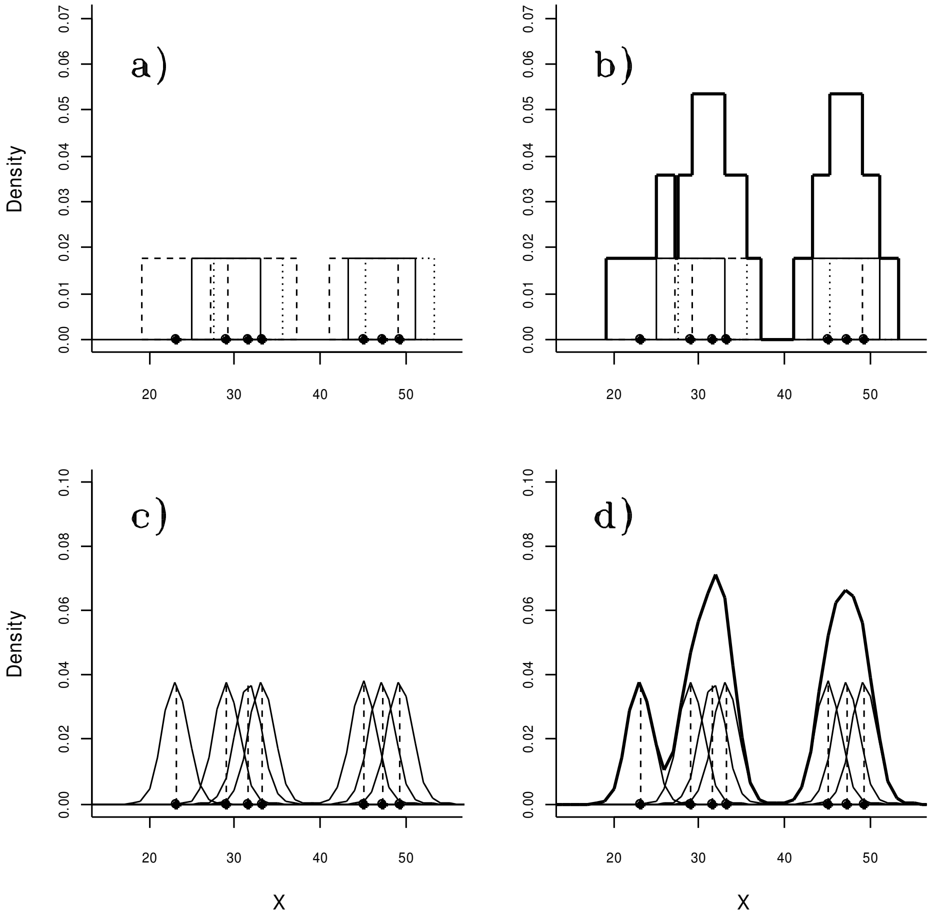


Fig. 2. Estimation of the density of one single variable with a set of 7 occurrences. a) The naïve estimator places a box of height $1/(2nh)$ and width $2h$ on each observation, where n is the number of observations and h a smoothing parameter chosen by the user. b) the summation of these boxes results in an estimation of the density function with a ragged aspect (bold curve). c) Kernel smoothing places a “bump” of area $1/n$ on each observation. d) The sum of these bumps lead to a smooth estimation of the density of the variable (bold curve).

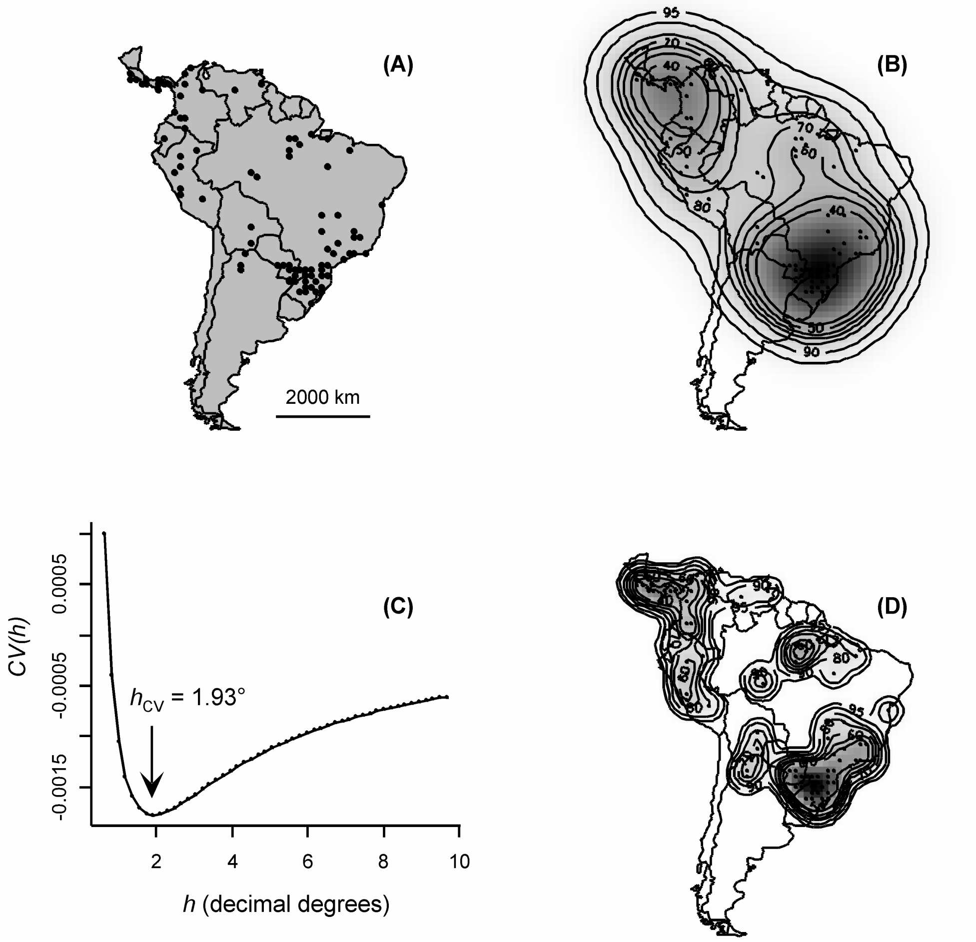


Fig. 3. Estimation of the geographical distribution of *Cedrela fissilis* Vell. in South America. a) Distribution of the 121 occurrences in South America. The occurrence coordinates are stored in decimal degrees. b) Kernel estimation of the distribution, with the *ad hoc* choice of the smoothing parameter h_{opt} = 6.93° (see text for the formula). The contour lines indicate the boundaries of the distribution ranges at various levels. Darker areas correspond to areas with a higher probability density of collecting a specimen of the species. c) Choice of the smoothing parameter *h* by Least Square Cross Validation (LSCR). The Cross-validation criteria *CV*(*h*) is minimised for h_{CV} = 1.93°. d) Kernel

estimation of the species distribution, with a smoothing parameter equal to h_{CV}. The contour lines indicate the boundaries of the distribution range at various levels.

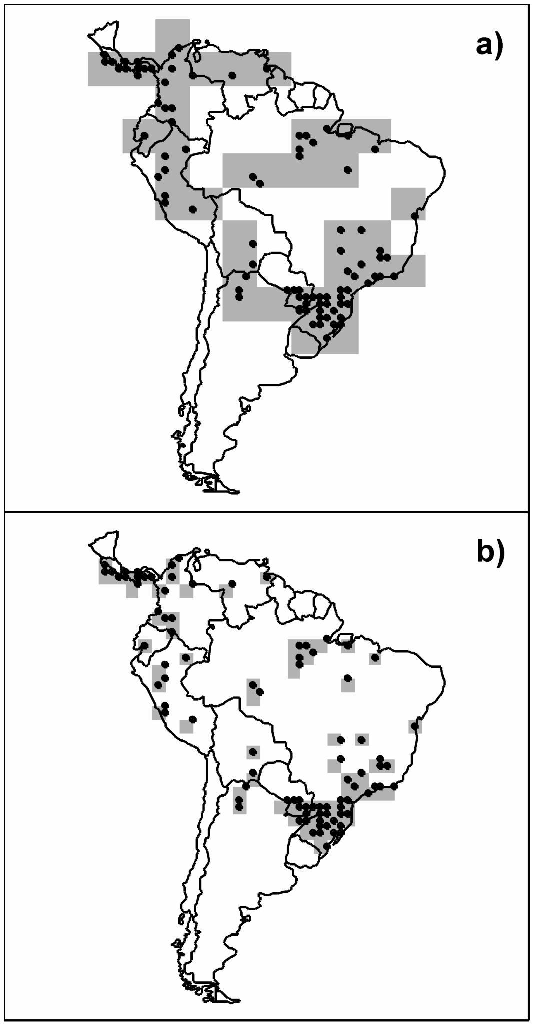


Fig. 4. Estimation of the Area Of Occurrence (AOO) for *Cedrela fissilis* using two size for the grid cells: a) AOO estimated with grid cells measuring 5° longitude × 5° latitude. b) AOO estimated with grid cells measuring 2° longitude × 2° latitude.

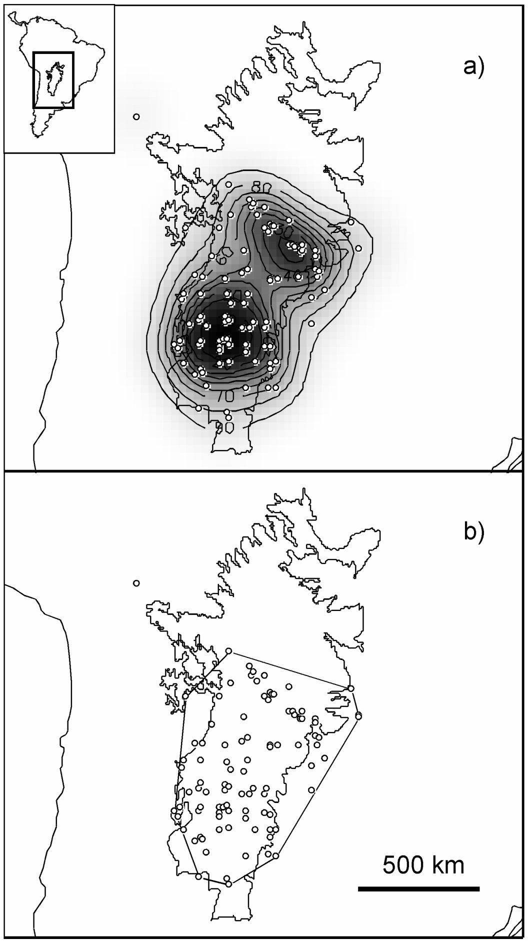


Fig. 5. Distribution pattern of the 48 occurrences of *Schinopsis quebracho-colorado* in South America. The limits of the Chaco region are displayed on each map of the South America. a) Estimation of the Geographical Distribution using the kernel estimator with a smoothing parameter equal to $h_{\text{opt}} = 1.05°$. The contour lines indicate the boundaries of the distribution range at various levels. b) Minimum convex polygon encompassing all occurrences of *S. quebracho-colorado* (one outlier has been excluded). The 48 occurrences of the species are also displayed.

Annexe 8

Article 8 :

“Efficiency of spreading maize in the garrigues to reduce wild boar damage to Mediterranean vineyards”

Clément Calenge, Daniel Maillard, Pascal Fournier & Carole Fouque.

European Journal of Wildlife Research (2004), 50, 112-120.

Eur J Wildl Res (2004) 50: 112–120
DOI 10.1007/s10344-004-0047-y

ORIGINAL PAPER

C. Calenge · D. Maillard · P. Fournier · C. Fouque

Efficiency of spreading maize in the garrigues to reduce wild boar (*Sus scrofa*) damage to Mediterranean vineyards

Received: 8 September 2003 / Accepted: 1 June 2004 / Published online: 9 July 2004

Abstract The aim of this work was to assess the efficiency of dissuasive spreading of maize to reduce the level of wild boar damage to vines in a Mediterranean area (Puéchabon, southern France). The 50 wine growers of Puéchabon were all questioned about the annual losses caused by the wild boar in each vineyard of the study area between 1990 and 1992. We also studied the distribution of the damage on a smaller scale, by mapping the damaged vines within two vineyards. In summer 1993, we distributed 4.7 tons of maize in the woods , and then questioned the wine growers about the losses in each vineyard for 1993. During 1990–1992, on both large and small scales, the damage had a patchy distribution, with more patches of damage occurring close to the woods. Moreover, the later the vines ripened, the less frequent the damage. The severity of the damage was only affected by the distance of the vineyard from woods. In total, the wild boar consumed 20,049 kg of grapes each year between 1990–1992 (193 kg/ha), and 63% of the vineyards were damaged. In 1993, both the density and the compensation paid in the département increased threefold. However, in our study area, both the proportion of damaged vineyards (36%) and the level of damage in the vineyards (151 kg/ha) were reduced, saving more than 60% in compensation. The dissuasive spreading of maize is therefore an efficient tool to reduce the level of damage to vineyards.

Keywords Compensation · Deterrent · Mediterranean habitat · Patchy distribution · Supplementary feeding

---

C. Calenge · D. Maillard (✉) · P. Fournier · C. Fouque
Office national de la chasse et de la faune sauvage,
95 rue Pierre Flourens, 34098
Montpellier Cedex 05, France
E-mail: d.maillard@oncfs.gouv.fr
Tel.: +33-467-107804
Fax: +33-467-107802

Introduction

During the last decades, the wild boar bag has steadily increased in France (37,500 wild boar were shot in 1973, 200,000 in 1993, and 343,700 in 1998), and in certain regions, crop damage is becoming increasingly unacceptable to the farmers (Vassant 1996). In the Mediterranean region in particular, the wild boar population has risen sharply (from 4,000 boars culled in 1973 to 62,000 in 1997). Compensation for wild boar damage to vineyards, the main local crop, has also increased (from €110,125 in 1978 to €463,340 in 1997). Because of the process of rural decline, more and more fields are left uncultivated, allowing a deeper penetration of the boars in the cultivated fields. In addition, this problem can only worsen with the gradual shrinkage of French viticultural areas increasingly planted with high quality vines. The concern of wine growers is therefore growing year by year.

Three-quarters of the wild boar's food intake is made up of two or three main items (Leranoz-Urtasum 1983; Fournier-Chambrillon et al. 1996). These staples always belong to the most abundant and most easily accessible food source for the period of the year. In the Mediterranean region, the abundant crop of acorns produced by the holly oaks (*Quercus ilex*) is the wild boar's main source of food from September to June. However, acorns on the ground do not keep long enough for the animals to subsist on them from one year to the next (Fournier-Chambrillon et al. 1996), and they are no longer available in summer. During this period, other natural foods are also limited (Fournier-Chambrillon et al. 1994a). In late summer, grapes are the only edible material available in large quantities in our study area.

There is therefore a considerable need for efficient management tools to reduce the level of damage to crops. Numerous methods have been developed to effectively limit such damage. Methods based on the creation of acoustic, olfactive or gustative aversions have not given satisfactory long-term results (Vassant and Boisaubert 1984; Vassant 1994a). On the other hand, electric fences are an effective means to prevent access to crops for wild boars and may be recommended for any type of crops (Boisaubert et al. 1983; Vassant and Boisaubert 1984). However, continuous surveillance is required to prevent the deterioration of both fences and batteries. In addition, electric fences may be efficient to solve local problems, but their large-scale use is problematic.

The most effective method for reducing the level of damage on crops seems to be the dissuasive spreading of maize. Indeed, this method has proved to be an efficient means for the protection of cereal crops (oats, wheat, etc.) in several areas (Vassant and Breton 1986; Vassant et al. 1992; Vassant 1994a, 1994b; Geisser 1998). In other countries in Europe, this method is even regulated by law (e.g. in Wallonie, Belgium, by ministerial decree published 13 June 2003). On the other hand, the efficiency of dissuasive spreading of grain to protect the vineyards still has to be fully tested. In fact, we do not know whether maize spreading in the garrigue scrublands during the warmest period of the year will prevent wild boar from going to the vineyards to feed on the sweet, water-rich fruit.

A good understanding of the factors that determine both the occurrence and the severity of damage is an essential step toward the development of this technique in a Mediterranean habitat (Geisser 1998). According to the literature, the main factors affecting the level of damage to crops are the species cultivated (Mackin 1970; Andrzejewski and Jezierski 1978; Merriggi and Sacchi 1992; Genov et al. 1995; Onida et al. 1995; Geisser 1998), the distance of the crops from woods (Merriggi and Sacchi 1992; Genov et al. 1995; Onida et al. 1995), the ripening period of the crop (Vassant and Breton 1986; Vassant et al. 1992; Lavoisier et al. 1996), the density of the wild-boar population (Andrzejewski and Jezierski 1978; Vassant 1994a; Geisser 1998; Spitz and Lek 1999), and the availability of natural food in the woods (Mackin 1970; Andrzejewski and Jezierski 1978; Genov et al. 1995).

In this study, we investigated the spatial pattern of the damage on both a large scale (distribution of the damage over the study area) and a small scale (distribution within two vineyards), because many authors have stressed the fact that resource selection by the animals may differ on different spatial scales (Johnson 1980; Otis 1997, 1998). Our aim was to identify the factors affecting both the occurrence and the severity of wild boar damage to crops, paying special attention to the ripening period of the grapes and to the distance of the crops from woods. The efficiency of dissuasive spreading of maize to limit wild boar frequentation of the vineyards was then tested, and finally we estimated the financial benefits of this type of operation.

Materials and methods

Study area

The study area is located in southeastern France, in the Hérault département (French district covering 615,000 ha), 35 km northwest of Montpellier, in the Puéchabon commune (latitude: 43°40′N, longitude: 3°36′E). This commune, covering 3,125 ha, is situated at the junction of the large, wine-growing plain of Béziers and the garrigue scrublands of Montpellier. The plain stretches far out into the wooded mountains over a width of 2–5 km. The wild boars frequently cross this plain when they pass from one plateau to another, venturing into the vineyards to feed when the grapes are ripe (Fournier-Chambrillon et al. 1994b). Between 1990 and 1992, the hunting bag was 0.7 animals culled per km^2. The hunting bag increased to 2 animals culled per km^2 in 1993, because of a management policy aimed at increasing the wild boar population over the whole Hérault département (limitation of both the number of culled adult females and the duration of the hunting season).

The garrigue (covering 3,000 ha), made up of low woody plant communities of typical Mediterranean species (*Phillirea latifolia* L., *Pistacia lentiscus* L., *Pistacia terebinthus* L., *Buxus sempervirens* L., *Juniperus oxycedrus* L., *Quercus coccifera* L.), is interspersed by coppices of holly oak (*Quercus ilex* L.) with some Aleppo pines (*Pinus halepensis* Mill.) and pubescent oak (*Quercus pubescens* Willd.). Acorns produced by the holly oak are the main food source of the wild boar in fall and winter (Fournier-Chambrillon et al. 1994b). Maillard (1996) estimated that the mast crop on the commune varied roughly between 300–500 kg ha^{-1}, depending on the tree density. The overall mast crop each year on the whole commune varies therefore between 900–1,500 tons. However, these acorns are no longer available in summer (most of them are eaten or rotten; Maillard 1996).

In our study area, the vineyard covers 125 ha, i.e. 45% of the farmed land, the rest being made up of fallows, fodder crops and olive groves which, for the greater part, have been left unattended. The crops found in the study area are representative of the region. Here, one finds the traditional varieties of vines that are used to make the wine from the hillsides of Languedoc. Within the study area, 283 vineyards were counted. Grapes are the most important crop, with very high yields. The different varieties of grape ripen at different times, which may also vary from year to year. The early varieties ripen between mid-July and mid-August and are not very common (2.3% of the total area planted with vines). Others ripen between mid-August and mid-September; they represent 34.4% of the vineyards. Finally, there are varieties which reach maturity between mid-September and mid-October. They are the most widespread and cover 63.3% of the vineyards.

Survey of wild boar damage to vineyards

Distribution of damage within the study area

To assess the impact of wild boar damage to the vineyards, a survey was carried out among all the wine growers in the study area, in 1992. The 50 vineyard owners were questioned regarding the degree (estimated in kilograms) of damage they had suffered in their vineyards from 1990–1992. According to the wine growers, the distribution of the damage in our study area did not change between 1990 and 1992, and the compensation paid was similar for these 3 years (between €6,000 and €6,300). Thus, they did not inform us on the level of damage to vineyards each year, but rather gave us an average annual estimation of the losses in each vineyard over this period. The level of damage in each vineyard was then mapped using Arcview GIS (ESRI 1996). The vineyards for which the estimated losses were higher than 300 kg ha^{-1} were considered to be severely damaged.

We computed the distance from woods for each vineyard. Five distance classes were defined:

1. contiguous with the woods,
2. not contiguous with the woods but < 50 m from the woods' edge,
3. between 50 and 100 m,
4. between 100 and 400 m and
5. further than 400 m from the woods.

The varieties of vines planted in each vineyard were classified into three categories, according to their ripening period:

1. the "early varieties", which ripen between mid-July and mid-August,
2. the "normal varieties", which ripen between mid-August and mid-September, and
3. the "late varieties", which ripen between mid-September and mid-October (Fig. 1).

We tested the main effects of distance from the woods and ripening period, as well as interactions between these factors, on the probability of damage occurrence using the generalized linear model, with a logit link (Mc Cullagh and Nelder 1989). We repeated the same analyses for the occurrence of severe damage, restricting them to damaged vineyards. However, due to the low number of damaged vineyards, we could not test the effect of interactions between distance from woods and ripening period on the frequency of severe damage. We therefore considered that these two factors were additive. We used the R software (Ihaka and Gentleman 1996) to perform these analyses.

Finally, we investigated the spatial pattern of damage frequency on the study area. We considered whether the

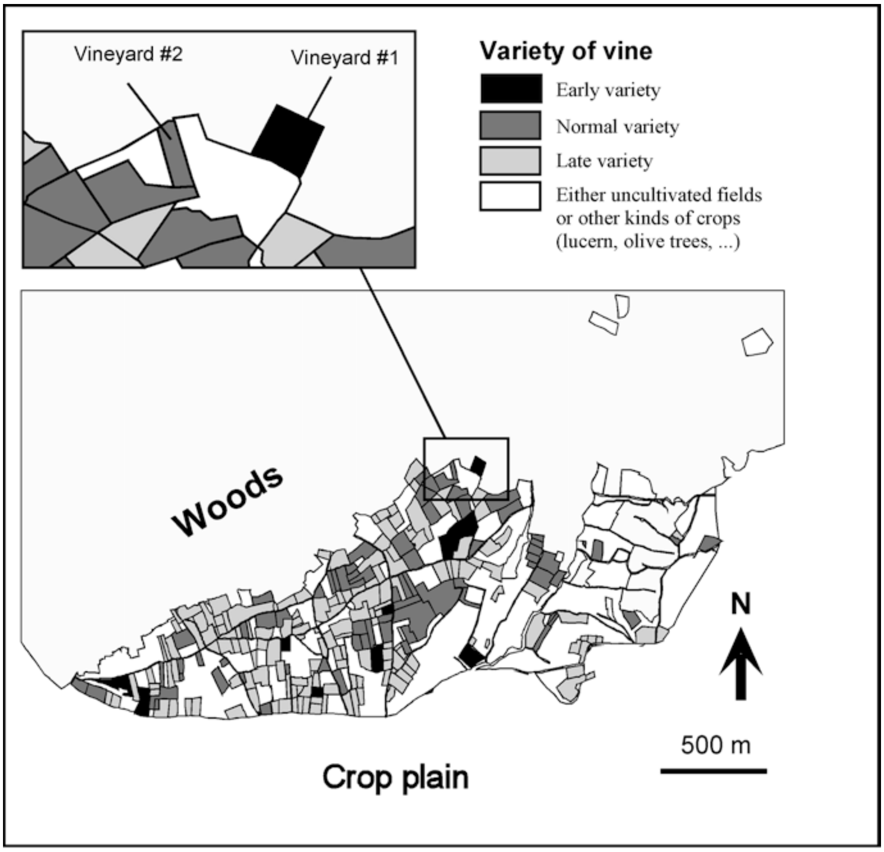


Fig. 1 Time of ripening of the vines planted in the vineyards of Puéchabon (southern France). In *vineyards #1* and *#2* each damaged vine was recorded on the map. The early varieties ripen between mid-July and mid-August, the normal varieties, between mid-August and mid-September and the late varieties, between mid-September and mid-October

frequency of wild boar damage had a clumped distribution; the existence of a possible spatial auto-correlation was tested with Geary's test (Chessel et al. 1997). Let x be the expected frequency of either damage or severe damage. Two vineyards having a common edge were considered as neighbours. The variance of x was rewritten using the following formula:

$$Var(x) = \frac{1}{2} \sum_{(i)\ \text{neighbour(j)}}^{n} \frac{1}{n^2}(x_i - x_j)^2 + \sum_{(i)\ \text{not neighbour (j)}}^{n} \frac{1}{n^2}(x_i - x_j)^2$$

Thus, the "total" variance of x may be considered as the sum of a "local" component (i neighbour of j) and a "global" component (i not neighbour of j). Let c be the ratio between the local variance and the total variance. Under the assumption that no spatial pattern of x occurs, c should be equal to 1. A ratio significantly lower than 1 indicated a clumped distribution of x (i.e. the local variance is lower than the total variance). The equality of c to 1 was tested using a permutation test (Chessel et al. 1997).

Distribution of the damage within two vineyards

We investigated the distribution of the damage within two vineyards of our study area (Fig. 1). Vineyard #1 was square-shaped, with three edges contiguous with the woods, and was planted with an early variety of vine. Vineyard #2 was rectangular, with only one edge contiguous with the woods, and was planted with a "normal" variety of vine, ripening much later. The east and west edges of this field were contiguous with uncultivated fields with dense herbaceous cover. In each vineyard, we noted the percentage of each vine damaged by the wild boar. We defined six ordered classes to describe damage:

1. 0% of the grapes consumed by wild boar,
2. < 10%,
3. 10–25%,
4. 25–50%,
5. 50–75%, and
6. > 75% of the grapes lost.

Within each vineyard, we considered whether the frequency of wild boar damage was clumped in distribution, using a Geary test. We then used the lowess regression to smooth the damage distribution within the vineyard (Chessel et al. 1997). We performed this regression on the number of neighbours that minimised the lowess error. These analyses were carried out using the ADE-4 software (Thioulouse et al. 1997). Finally, we tested the correlation between the distance from woods and the level of damage using Spearman coefficients.

Dissuasive spreading of maize

In 1993, 4.7 tons of grain maize were distributed in the woods between 28 July and 6 September, forming a 4.5 km trail situated between 500 m and 1 km from the vineyards (an average amount of 25 kg day^{-1} km^{-1}). The maize was poured in a steady sideways flow from a tractor-mounted tank. The daily distribution took on average 3 h. No other deterrent (e.g. electric fences) was applied in our study area during the focus period (1990–1993). To assess the efficiency of this technique to reduce the damage, we again questioned the 50 wine-growers in the study area on the damage (estimated in kilograms) suffered that year. For each vineyard, we determined whether the damage in 1993 was reduced, increased or similar to previous years. We then modelled the probability that a vineyard was less damaged during the application of the deterrence method using a generalised linear model with a logit link, restricting these analyses to the vineyards damaged during previous years. As for the models of severe damage, we considered that the effects of distance from woods and time of ripening of the vines were additive. We used the R software (Ihaka and Gentleman 1996) to perform these analyses.

Results

Survey of wild boar damage to vineyards

Distribution of damage within the study area

In our study area, wild boar damaged 43.8% of the vineyards (124 vineyards out of 283) between 1990 and 1992, and on average, they consumed 193 kg grapes ha^{-1} (SE = 29). About 3.5% of the grape production was lost due to this predation. The probability of both damage and severe damage showed a highly clumped distribution (Fig. 2, damage: $c = 0.52$, $n = 287$, $P < 0.001$, severe damage: $c = 0.63$, $n = 287$, $P < 0.001$). In addition, the frequency and the severity of the damage varied according to the distance from the woods (frequency: $\chi^2 = 25.6$, $df = 4$, $P < 0.001$, severity: $\chi^2 = 18.4$, $df = 4$, $P = 0.001$). The closer a vineyard was to the woods, the higher the frequency of damage (Fig. 3A). The same kind of pattern was observed for the frequency of severe damage, though the probability of severe damage was less in the vineyards contiguous with the woods than in the vineyards not contiguous with the woods but < 50 m from the woods' edge (Fig. 3B).

The frequency of damage was also affected by the ripening period of the varieties cultivated ($\chi^2 = 15$, $df = 2$, $P < 0.001$). The earlier a vine reached maturity, the higher the frequency of damage. On the other hand, we did not identify any effect of the ripening period on the frequency of severe damage ($\chi^2 = 3.4$, $df = 2$, $P = 0.18$).

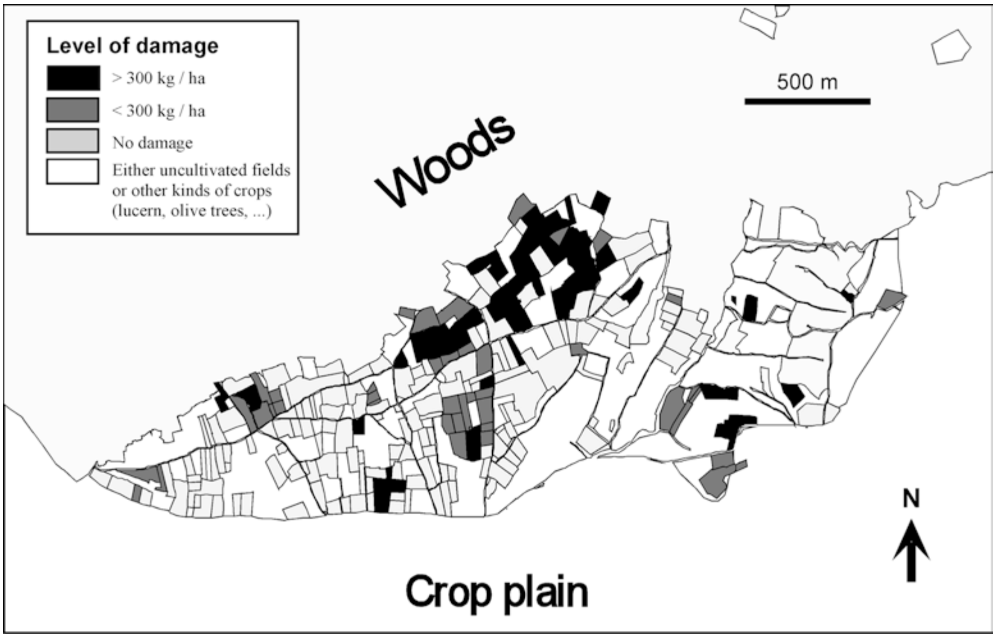


Fig. 2 Distribution of the wild boar damage in the vineyards of Puéchabon (southern France) during 1990–1992, without supplementary feeding

We found no effect of the interactions between the ripening period and the distance from the woods on the probability of damage ($\chi^2 = 9.9$, $df = 7$, $P = 0.19$).

Distribution of the damage within two vineyards

We examined 274 vines in vineyard #1, and 448 in vineyard #2. Vineyard #1, planted with an early variety, was more severely damaged (47.4% of the vines showing the presence of wild boar damage, i.e. 130 vines) than vineyard #2 (15.4%, i.e. 69 vines), planted with a normal variety. In both areas, the damage showed a highly clumped distribution (Fig. 4, vineyard #1: $c = 0.27$, $n = 274$, $P < 0.001$, vineyard #2: $c = 0.62$, $n = 448$, $P < 0.001$). We found a significant correlation between the distance from the woods and the intensity of the damage on a vine in both areas (vineyard #1: $R_S = -0.70$, $P < 0.001$, vineyard #2: $R_S = -0.22$, $P < 0.001$). Within the two vineyards, the closer a vine was to the woods, the heavier the damage (Fig. 4). Moreover, in vineyard #2, the vines located close to the uncultivated fields, at the east and west edges of the vineyard, were more severely damaged than those located towards the centre of the vineyard.

Efficiency of dissuasive spreading of maize to reduce damage

As a whole, both the frequency and the intensity of damage were slightly reduced during the application of the deterrence method in 1993 (Table 1). All the vineyards where no damage was recorded in 1990–1992 remained undamaged in 1993. The dissuasive spreading of grain maize resulted in a small decrease of the mean amount of estimated losses per surface area, although it was not significant ($t = 1.77$, $df = 287$, $P = 0.08$). The level of the damage decreased in 35% of the vineyards damaged in 1990–1992, and increased in only 7% of the cases. The damage did not change between the two periods in most vineyards (58%).

The probability that a vineyard damaged in 1990–1992 was less damaged in 1993 was not affected by the distance from woods ($\chi^2 = 6.65$, $df = 4$, $P = 0.16$) or by the ripening period of the vines ($\chi^2 = 4.2$, $df = 2$, $P < 0.12$), and seemed to occur randomly throughout the study area (Fig. 5).

Discussion

In 1990–1992, both the proportion of vineyards damaged by wild boar and the intensity of the damage increased with the proximity to woods. We also identified a clumped distribution of the damage. In addition, varieties of vine that ripen in early summer were more frequently damaged than others. The dissuasive spreading of grain maize resulted in a slight decrease in both the proportion of damaged vineyards and the intensity of the damage in our study area.

Wild boar preferentially wander along known paths for foraging, causing damage in the neighbourhood of selected foraging places, resulting in a patchy distribution of the damage (Kristiansson 1985; Welander 2000). Thus, even if the losses were rather low on the scale of the study area (3.5% of the grape production), locally, certain vineyards were strongly damaged, and certain wine growers were more affected by the losses than others. On both large and small scales, the patches of damage occurred mainly close to the woods. Security seems to be a factor of the utmost importance for the wild boar (Spitz and Janeau 1995). The animals will use

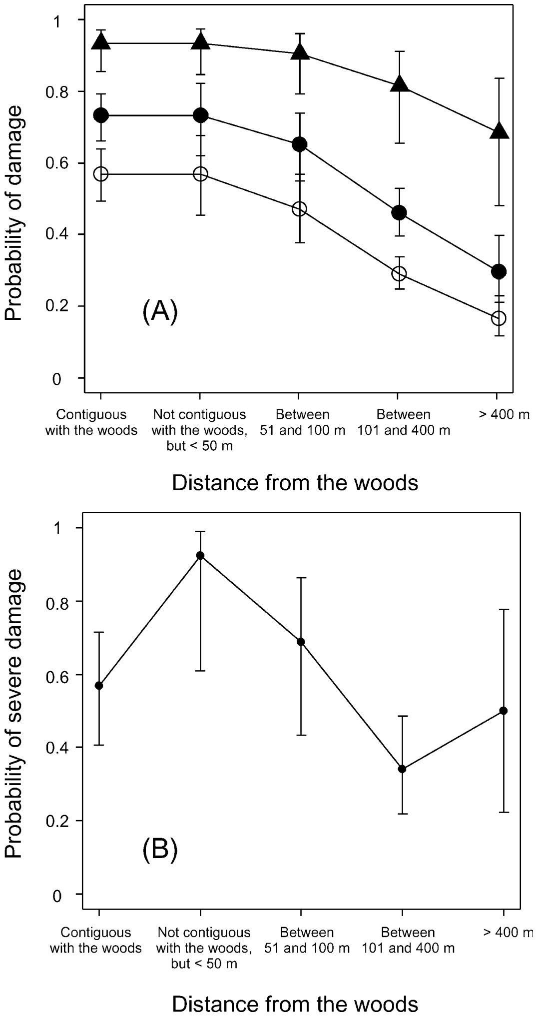


Fig. 3 **A** Logistic model of the occurrence of wild boar damage in the vineyards according to the distance from woods and the time of ripening of the variety planted (*filled triangles*: variety ripening between mid-July and mid-August; *filled circles*: variety ripening between mid-August and mid-September; *empty circles:* variety ripening between mid-September and mid-October). **B** Logistic model of the probability of severe damage in the damaged vineyards, according to the distance from woods (±95% confidence intervals of the estimations)

the vineyards while taking care to stay close to the field edges, allowing easier escape in case of threat. This kind of behaviour has also been noted in other areas (Vassant and Breton 1986; Gérard and Campan 1988; Merriggi and Sacchi 1992; Genov et al. 1995; Onida et al. 1995; Spitz and Lek 1999).

As long as sufficient food is available in the woods, the wild boar scarcely feed on crops (Andrzejewski and Jezierski 1978; Mackin 1970; Genov et al. 1995). In our study area, there is a constant increase in the availability of food in the woods from July to October (Fournier-Chambrillon et al. 1996). The summer (July–August) is a period of food shortage, and the wild boar feed frequently on vines. From September, there is a gradual increase in acorn availability, and consequently, a decrease in the frequency of the damage to vines. Thus, the later a vine ripens, the less frequent the damage.

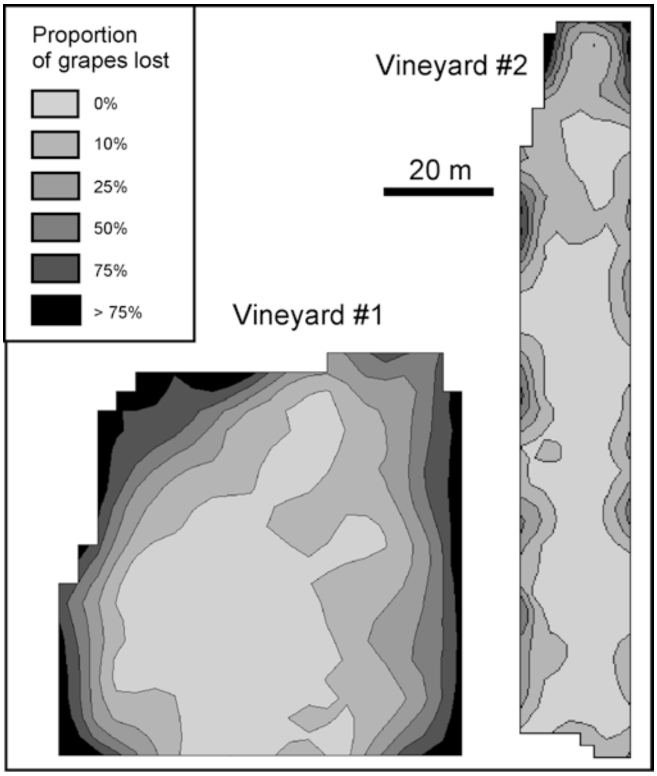


Fig. 4 Smoothed representation of the proportion of grapes consumed by the wild boar on each vine: **A** within vineyard #1 (see Fig. 1; lowess regression on 12 neighbours); **B** within vineyard #2 (see Fig. 1; lowess regression on 14 neighbours)

The dissuasive spreading of maize results in an increase of the food availability in the woods, and many authors have stressed its efficiency to reduce the damage to cereal crops (Vassant and Breton 1986; Vassant et al. 1992; Vassant 1994a, 1994b; Geisser 1998). This method also seems efficient to reduce damage to vines. In our study area, the hunting bag increased threefold between 1990-1992 (wild boar culled, 0.7/km^2) and 1993 (wild boar culled, 2/km^2) in the woods of Puéchabon, whereas the hunting pressure did not change between the two periods (similar number of hunters, number of days hunted per year and numbers of hounds), suggesting an increase in the wild boar population in 1993. This augmentation was also recorded for the whole Hérault

Table 1 Comparison of the level of damage to vineyards without (1990–1992) and with (1993) dissuasive spreading of maize in the vineyards of Puéchabon (south of France)

	1990–1992	1993
Total loss (kg)	20,049	15,716
Losses (kg/ha)	193 (SE = 29)	151 (SE = 21)
No. of damaged vineyards	124 (43%[a])	105 (36%[a])
Compensation paid (€)	6203	4797

[a] Proportion of vineyards damaged by wild boar

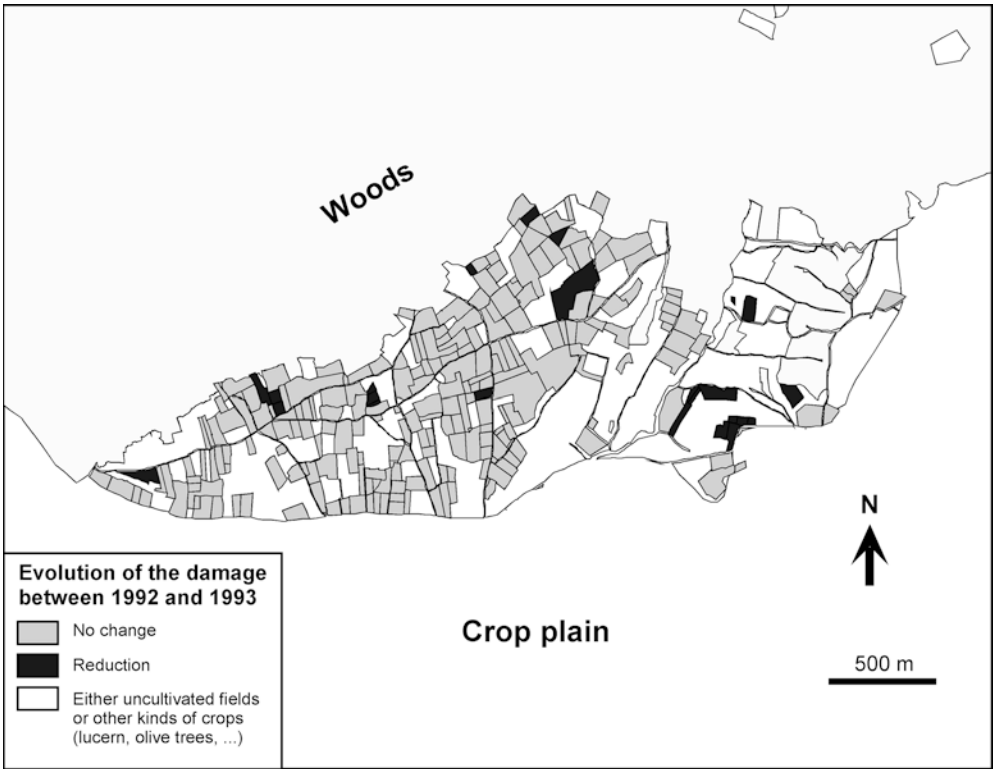


Fig. 5 Changes of the wild boar damage to the vineyards of Puéchabon (southern France) between 1990–1992 (without supplementary feeding) and 1993 (dissuasive spreading of maize in the woods). The vineyards for which the damage increased between the two periods were both scarce (7%) and small in area (total = 2.7 ha), and are not displayed

département (from 3,500 to 5,500 wild boar culled), which led to a considerable increase in the level of damage to crops: compensation paid yearly on the whole département increased from €24,400 in 1990–1992 to €544 000 in 1993. However, in our study area, we observed a slight decrease in both the level of damage and the number of damaged vineyards in 1993. Although we acknowledge that a study of the evolution of the availability of natural food in the woods between 1990–1992 and 1993 would have strengthened any conclusion, this decrease is certainly the consequence of the dissuasive spreading of grain. This therefore emphasizes the efficiency of this measure of protection.

It is essential here to differentiate between the dissuasive and "attractive" food supply. In several areas, hunters spread maize throughout the year to attract the boars to their hunting territories. This additional food may result in an increase in population size and lead to a long-term increase in the damage (e.g. Andrzejewski and Jezierski 1978). We strongly warn here against this practice, and we stress that the dissuasive spreading of maize is different both in its aim and results. When used as a deterrent, the maize is spread only over a very short period, as long as the grapes are ripe. The amount of maize spread (~5 tons) is negligible in comparison to the amount of acorns available in the area one month later (estimated at 900–1,500 tons over the whole Puechabon commune; Maillard 1996). The effect of this deterrence method on the population size is therefore minor. Indeed, on our study area, dissuasive maize-spreading has been continued for 10 years after the study. The hunting bag (a rough index of the population size) randomly fluctuated during this period, despite the long-term application of the method (Fig. 6). Because this spreading resulted in a considerably lower damage to the vines (Table 1), we stress that the dissuasive spreading of maize is a replacement food source rather than an additional food source.

Financial gain

The fact that the dissuasive spreading of maize lead to a decrease in the level of damage was expected. Actually, we need an estimation of the financial gain of such a management tool to inform land managers on the efficiency of this method. The current cost of maize spreading may be estimated at €3,461 (4.7 tons of grain maize: €1,433, equipment for the distribution of maize: €381, running expenses for a vehicle for 40 days × 15 km: €183, wages for 40 days × 3 h: €1,464).

In 1993, when the deterrence measure was applied, damage compensation paid amounted to €4,797. To draw up a complete economic balance-sheet, we also estimated the cost of damage compensation that would have been paid in our study area in the absence of deterrence measures. In fact, in 1993, damage to the vineyards in the département increased 3.4-fold compared to the three preceding years. A similar trend was recorded in various sectors of the same département surrounding our study area. So, to reach an estimate of the expenses which, according to our study, should have

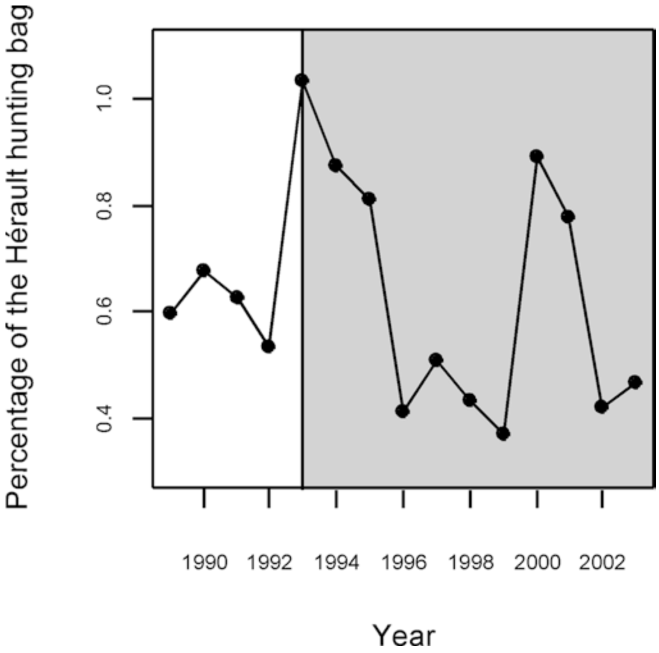


Fig. 6 Fluctuations in the wild boar hunting bag at Puéchabon (district of Hérault, France) from 1989–2003. The hunting bag is measured as the percentage of the whole hunting bag of Hérault. From 1993–2003, the dissuasive spreading of maize was used to reduce the amount of the damage to vines (*grey period*)

been incurred, we multiplied the expenses for wild boar damage in the study area in the years 1990–1992 (estimated at €6203) by 3.4, i.e. a total of €21,090.

The final balance is positive: it shows savings of more than €12,800, i.e. a decrease of 61% in the compensation paid:

- Estimated losses without supplementary food: €21,090
- Estimated losses with supplementary food: €4,797
- Cost of dissuasion: €3,461
- Net final balance: €12,832

In addition to the financial gains, a further positive aspect is that the amount of damage went down. This satisfies the farmers and thus minimises sources of conflict between farmers and hunters.

Management implications

After this study, dissuasive grain-spreading has been continued by the hunters in Puéchabon, also around some isolated vineyards in the forested hill sides. Every year, the people concerned meet in June, and devise a plan of action for the next protection campaign based on the mast crop and last year's hunting bag, on the understanding that the hunters are responsible for the daily distribution of the maize and the erection of crop fences and the farmers ensure their upkeep. This action was completed by the construction of a few electric fences to protect the most sensitive vineyards, i.e. the vineyards planted with early variety of vines and close to the woods. In spite of a constant increase in wild boar numbers, their pressure on the vineyards is negligible and a climate of good understanding has been reinstated between wine growers and hunters (Maillard and Sanier 1996).

We emphasise that maize should not be provided at localised points. Rather, supplying the maize over 5–10 m wide strips guarantees that the supplementary food is accessible to all wild boar and that one male or a group of females does not drive the other wild boar off (Vassant et al. 1992). The maize should be supplied as long as the crops are ripe, at a rate of 2–4 kg/100 m, according to the recommendations of Vassant et al. (1992). The managers should acknowledge that the dissuasive spreading of maize is just a palliative in the protection of crops and is unlikely to be as efficient in cases of overdense population. Thus, this measure should be added to other measures intending to control the population density, as part of an effective policy of management of wild boar populations.

Acknowledgements We are grateful to the wine growers of Puéchabon who answered our questions, as well as to the many students who helped during data collection. We also warmly thank Eric Baubet (Office national de la chasse et de la faune sauvage) for his useful comments on an earlier draft of this paper. The experiments carried out in this study comply with the current laws of France.

References

Andrzejewski R, Jezierski W (1978) Management of a wild boar population and its effects on commercial land. Acta Theriol 23:309–339

Boisaubert B, Bouldoire JL, Vassant J (1983) Bilan de cinq années d'expérimentation de protection des cultures par clôtures électiques dans le département de la Haute-Marne. Bull Mens Off Natl Chasse 68:15–26

Chessel D, Thioulouse J, Champely S (1997) Autocorrélation et composantes cartographiables. Documentation thématique du logiciel ADE-4. Fiche thématique 7.3. Villeurbanne

ESRI (Environmental Systems Research Institute) (1996) Using Arcview GIS. The geographic information system for everyone. Environmental Systems Research Institute, Redlands, CA

Fournier-Chambrillon C, Maillard D, Fournier P (1994a) Diet of the wild boar (*Sus scrofa* L.) inhabiting the Montpellier garrigue. IBEX JME 3:179–179

Fournier-Chambrillon C, Maillard D, Fournier P (1994b) Le régime alimentaire du sanglier en milieu méditerranéen. Approche des dégâts sur vigne. Bull Mens Off Natl Chasse 191:36-47

Fournier-Chambrillon C, Maillard D, Fournier P (1996) Variabilité du régime alimentaire du sanglier (*Sus scrofa* L.) dans les garrigues de Montpellier (Hérault). Gibier Faune Sauvage 13:1457–1476

Geisser H (1998) The wild boar (*Sus scrofa*) in the Thurgau (Northeastern Switzerland): population status, damage and the influence of supplementary feeding on damage frequency. Gibier Faune Sauvage 15:547–554

Genov PV, Tonini L, Massei G (1995) Crop damage by wild ungulates in a Mediterranean area. In: Botev N (ed) The game and the man, Proc IUGB, Sofia,pp 214–215

Gérard JF, Campan R (1988) Variabilité éco-éthologique chez le sanglier européen: comparaison des travaux français. Cah Ethol Appl 8:63–130

Ihaka R, Gentleman R (1996) R: A language for data analysis and graphics. J Comput Graph Stat 5:299–314

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

Kristiansson H (1985) Crop damage by wild boars in central Sweden. In: de Crombrugghe, SA (ed.) Proc 17th Congr IUGB Brussels, pp 605–609

Lavoisier E, Léon F, Havet P, Granval P (1996) Intérêt de la jachère pour réduire les dégâts de grand gibier dans les récoltes. Bull Mens Off Natl Chasse 214:51–55

Leranoz-Urtasum I (1983) Sobre la relacion des jabali (*Sus scrofa* L.) con la agricultura, en Navarra septentrional. Actas XV Congr Int Fauna Cinegetica Silvestre, Trujillo, Spain, pp 639–645

Mackin R (1970) Dynamics of damage caused by wild boar to different agricultural crops. Acta Theriol 15:447–458

McCullagh P, Nelder JA (1989) Generalized linear models, 2nd edn. Chapman and Hall, London

Maillard D (1996) Occupation et utilisation de la garrigue et du vignoble méditerranéens par le sanglier (*Sus scrofa* L.). Doctoral Thesis, University of Aix-Marseille III, France

Maillard D, Sanier M (1996) Comportement alimentaire, dégâts dans les vignes et protection. La Chasse dans l'Hérault 24:21–21

Meriggi A, Sacchi O (1992) Factors affecting damage by wild boars to cereal fields in northern Italy. In: Spitz F, Janeau G, Gonzalez G, Aulagnier S (eds) Proc Int Symp Ongulés/Ungulates 1991, SFEPM-IRGM, Paris, pp 439–441

Onida P, Garau F, Cossu S (1995) Damage caused to crops by wild boars (*S. scrofa meridionalis*) in Sardinia (Italy). IBEX JME 3:230–235

Otis DL (1997) Analysis of habitat selection studies with multiple patches within cover types. J Wildl Manage 61:1016–1022

Otis DL (1998) Analysis of the influence of spatial pattern in habitat selection studies. J Agric Biol Environ Stat 3:254–267

Spitz F, Janeau G (1995) Daily selection of habitat in wild boar (*Sus scrofa*). J Zool (Lond) 237:423–434

Spitz F, Lek S (1999) Environmental impact prediction using neural network modelling. An example in wildlife damage. J Appl Ecol 36:317–326

Thioulouse J, Chessel D, Doledec S, Olivier JM (1997) ADE-4 : a multivariate analysis and graphical display software. Stat Comput 7:75–83

Vassant J (1994a) Les techniques de prévention des dégâts de sanglier. Bull Mens Off Natl Chasse 191:90–93

Vassant J (1994b) L'agrainage dissuasif : résultats d'expériences. Bull Mens Off Natl Chasse 191:101–105

Vassant J (1996) Evolution of wild boar populations and damage to crops in France between 1978 and 1993. In: C.I.C. (ed.) Proc Schwarzwild Symp, Sopron, Ungarn, pp 131–141

Vassant J, Boisaubert B (1984) Bilan des expérimentations entreprises en Haute-Marne pour réduire les dégâts de sangliers (*Sus scrofa*) à l'encontre des cultures agricoles. In: Spitz F, Pépin D (eds) Proc Symp Int Sanglier. Colloques de l'INRA 22:187–199

Vassant J, Breton D (1986) Essai de réduction des dégâts de sangliers (*Sus scrofa scrofa*) sur blé (*Triticum sativum*) au stade laiteux par distribution de maïs (*Zea mais*) en forêt. Gibier Faune Sauvage 3:83–95

Vassant J, Jullien JM, Brandt S (1992) Reducing wild boar damage to wheat and oats in summer: study of the effectiveness of maize distribution in the forest. In: Bobek B, Perzanowski K, Regelin WL (eds) Proc 18th IUGB Congr, Global trends in wildlife management, Jagdellonian University, Krakow

Welander J (2000) Spatial and temporal dynamics of a disturbance regime. Wild boar *Sus scrofa* rooting and its effects on plant species diversity. Doctoral Thesis, Swedish University of Agricultural Sciences, Utgivningsort

Annexe 9

Equivalence entre deux formulations du problème statistique posé par l'ENFA

Equivalence entre deux formulations du problème statistique posé par l'ENFA

➔ Soit **Z** la matrice décrivant la valeur de *P* variables environnementales (colonnes) dans *N* pixels (lignes). La matrice **Z** est centrée pour la pondération uniforme. Cette matrice définit un nuage de *N* points dans un espace à *P* dimensions.

➔ A chacun des points de ce nuage est associé un poids d'utilisation qui indique la fréquence d'utilisation du pixel par l'espèce étudiée. Ces poids sont stockés dans le vecteur **p**. La somme des poids sur tous les pixels est égale à 1, donc :

$$\mathbf{p}^t\mathbf{1}_N = 1$$

➔ Enfin, à chaque colonne de **Z** est associé une pondération. Ces poids sont stockés dans le vecteur **c**

➔ On définit les matrices suivantes :

$$\mathbf{D} = \mathrm{Diag}(\ 1 / N\ )_{i = 1 \dots N}$$

$$\mathbf{D_p} = \mathrm{Diag}(\ \mathbf{p}\ )$$

$$\mathbf{Q} = \mathrm{Diag}(\ \mathbf{c}\ )$$

➔ Enfin, on définit le vecteur de marginalité **m**, qui définit les coordonnées du centre de gravité du nuage des points utilisés dans l'espace ayant pour origine le barycentre des pixels disponibles :

$$\mathbf{m} = \mathbf{Z}^t\ \mathbf{D_p}\ \mathbf{1}_n$$

<u>Première formulation :</u> Le problème de l'ENFA correspond à la recherche d'un vecteur **u** qui remplit les conditions suivantes :

$$\text{Problème 1 : } \begin{cases} \mathbf{u}^t\mathbf{Q}\mathbf{u} = 1 & (1.1) \\ \mathbf{u}^t\mathbf{Q}\mathbf{m} = 0 & (1.2) \\ \mathbf{Z}\mathbf{Q}\mathbf{u} = \mathbf{y} & \\ R = \dfrac{\mathbf{y}^t\mathbf{D}\mathbf{y}}{\mathbf{y}^t\mathbf{D_p}\mathbf{y}}\ \ Max & (1.3) \end{cases}$$

La condition 1.1 implique que le vecteur **u** est normé pour la métrique **Q**.
La condition 1.2 implique que le vecteur **u** est non corrélé au vecteur de marginalité pour la métrique **Q**.

Comme le vecteur **u** est normé, **y** est la projection des lignes de **Z** sur **u**. **y** définit donc des scores pour les lignes de **Z**. La condition 1.3 implique que le rapport variance des scores disponibles sur variance des scores utilisés doit être maximum (on maximise ainsi la spécialisation).

Seconde formulation : l'ENFA peut aussi être vue comme la recherche d'un vecteur **w** respectant les conditions suivantes:

$$\text{Problème 2 : }\begin{cases}\mathbf{w}^t\mathbf{Qm}=0 & (2.1)\\ \mathbf{ZQw}=\mathbf{a} & \\ \mathbf{a}^t\mathbf{D_p a}=1 & (2.2)\\ \mathbf{a}^t\mathbf{Da}=Max & (2.3)\end{cases}$$

Les deux problèmes sont équivalents. En effet, il existe un lien entre le vecteur **u** solution du problème 1 et le vecteur **w** solution du problème 2 :

$$\mathbf{u}=\frac{\mathbf{w}}{\|\mathbf{w}\|_{\mathbf{Q}}} \qquad (3)$$

Démonstration de l'équivalence :

➔ *Si* **w** *est solution du problème 2, alors* **u** *est solution du problème 1:*

Si **w** est solution du problème 2, et que l'on pose

$$\mathbf{u}=\frac{\mathbf{w}}{\|\mathbf{w}\|_{\mathbf{Q}}}$$

Alors **u** est **Q-**normé, et la condition 1.1 est respectée.
Si **w** est solution du problème 2, alors la condition 2.1 est respectée, et donc la condition 1.2 l'est aussi.
Enfin, le rapport de la condition 1.3 peut être calculé :

$$R=\frac{\mathbf{y}^t\mathbf{Dy}}{\mathbf{y}^t\mathbf{D_p y}}=\frac{\mathbf{u}^t\mathbf{QZ}^t\mathbf{DZQu}}{\mathbf{u}^t\mathbf{QZ}^t\mathbf{D_p ZQu}}=\frac{\mathbf{w}^t\mathbf{QZ}^t\mathbf{DZQw}}{\mathbf{w}^t\mathbf{QZ}^t\mathbf{D_p ZQw}}=\frac{\mathbf{w}^t\mathbf{QZ}^t\mathbf{DZQw}}{\|\mathbf{a}\|^2_{D_p}}=\mathbf{a}^t\mathbf{Da}$$

La condition 1.3 est donc également respectée.

➔ *Si* **u** *est solution du problème 1, alors* **w** *est solution du problème 2:*

Si **u** est solution du problème 1, et que l'on pose

$$\alpha = \frac{1}{\|\mathbf{ZQu}\|_{\mathbf{D_p}}}$$

et

$$\mathbf{w} = \alpha\, \mathbf{u}$$

Alors, la condition 2.1 est nécessairement remplie par le vecteur **w**.

On peut calculer :

$$\mathbf{a}^t\,\mathbf{D_p a} = \mathbf{w}^t\,\mathbf{QZ}^t\mathbf{D_p ZQw} = \alpha^2\mathbf{u}^t\,\mathbf{QZ}^t\mathbf{D_p ZQu} = \frac{\mathbf{u}^t\,\mathbf{QZ}^t\mathbf{D_p ZQu}}{\mathbf{u}^t\,\mathbf{QZ}^t\mathbf{D_p ZQu}} = 1$$

Et la condition 2.2 est également respectée.
Enfin, on peut montrer que :

$$\mathbf{a}^t\,\mathbf{Da} = \mathbf{w}^t\,\mathbf{QZ}^t\mathbf{DZQw} = \alpha^2\mathbf{u}^t\,\mathbf{QZ}^t\mathbf{DZQu} = \frac{\mathbf{u}^t\,\mathbf{QZ}^t\mathbf{DZQu}}{\mathbf{u}^t\,\mathbf{QZ}^t\mathbf{D_p ZQu}} = R$$

Donc, si **u** est solution du problème 1, R est maximum, et par le fait, $\mathbf{a}^t\,\mathbf{Da}$ aussi. Donc la condition 2.3 est aussi respectée.

Les deux problèmes sont donc équivalents.
On cherchera à résoudre le problème 2 qui est plus simple.

Si on trouve un vecteur **w** solution du problème 2, le normer donnera un vecteur **u** qui résout le problème 1.

Annexe 10

Commandes du logiciel R utilisées pour construire les figures du chapitre 4

Commandes du logiciel R utilisées pour construire les figures du chapitre 4

Nous présentons ci-dessous les commandes que nous avons utilisées pour construire les figures du chapitre 4 avec le logiciel R. Les commentaires sont précédés d'un #

```
######################################################
##                                                  ##
##                                                  ##
##            Chargement des données                ##
##                                                  ##
##                                                  ##
######################################################


library(adehabitat)
data(lynxjura)
fac <- as.matrix(lynxjura$locs[,c("X","Y")])
set.seed(2312)
fac[,1] <- jitter(fac[,1], factor=3)
fac[,2] <- jitter(fac[,2], factor=3)




######################################################
##                                                  ##
##                                                  ##
##                Cartes diverses                   ##
##                                                  ##
##                                                  ##
######################################################


## Carte ponctuelle (Figure 5)
opar <- par(mar=c(0.1,0.1,0.1,0.1))
plot(fac, asp=1, pch=16)
par(opar)


## Création d'une grille de quadrats et dénombrement des
## occurrences dans chacun (Figure 6)
opar <- par(mar=c(0.1,0.1,0.1,0.1), mfrow=c(2,2))
a4 <- ascgen(fac, cellsize=8000)
image(a4, axes=F)
box()
a3 <- ascgen(fac, cellsize=2000)
image(a3, axes=F)
box()
a2 <- ascgen(fac, cellsize=1000)
image(a2, axes=F)
box()
a1 <- ascgen(fac, cellsize=500)
```

```
image(a1, axes=F)
box()
par(opar)


## Lissage par la méthode du noyau (Figure 9)
gr <- ascgen(fac, cellsize=200)
fac <- as.data.frame(fac)
a1 <- kernelUD(fac, h=5000, grid=gr)
a3 <- kernelUD(fac, h=2000, grid=gr)
a2 <- kernelUD(fac, h=1000, grid=gr)


opar <- par(mar=c(0.1,0.1,0.1,0.1), mfrow=c(2,2))
image(a1, axes=F, addp=F)
box()
image(a3, axes=F, addp=F)
box()
image(a2, axes=F, addp=F)
box()
image(a2[[1]]$UD, axes=F, col="white")
points(fac, pch=16)
box()
par(opar)


######################################################
##                                                  ##
##                                                  ##
##         Comparaison du semis avec la CSR         ##
##                                                  ##
##                                                  ##
######################################################


####### Calcul de la fonction F à l'aide de la
####### bibliothèque splancs

library(splancs)
pol <- bbox(fac)
np <- nrow(fac)

## Estime une grille de points
grille <- gridpts(pol, 1000)

## Estime la fonction F pour un vecteur de distances
FF <- Fhat(fac, grille, dis <- seq(0,5000,length = 100))

## Estime F théorique (pdense mesure la densité globale d'un pattern)
Ftheo <- Fzero(pdense(fac, pol), dis)


####### Calcul de la fonction G à l'aide de la
####### bibliothèque splancs

### Calcul de G
GG <- Ghat(fac, dis)


####### Simulations de la CSR
g <- list()
```

```
f <- list()
for (i in 1:99) {
  spts <- csr(pol, np)
  g[[i]] <- Ghat(spts, dis)
  f[[i]] <- Fhat(spts, grille, dis)
}
g <- do.call("rbind", g)
f <- do.call("rbind", f)
mg <- apply(g, 2, min)
Mg <- apply(g, 2, max)
mf <- apply(f, 2, min)
Mf <- apply(f, 2, max)


####### Courbe EDF de la fonction F (Figure 11)

x11(7,3.5)
par(mfrow=c(1,2), cex.lab=0.9,mar=c(5,5,1,1))
plot(dis,FF, ty="l", xlab="d (mètres)",ylab="F(d)", lwd=2)
lines(dis, mf, lty="dashed")
lines(dis, Mf, lty="dashed")
lines(dis, Ftheo)


plot(Ftheo, FF, ty="l", lwd=2, xlab=expression(paste(F[th], "(d)")),
     ylab=expression(paste(hat(F), "(d)")))
lines(c(0,1),c(0,1), lwd=1)
lines(Ftheo, mf, lty="dashed", lwd=1)
lines(Ftheo, Mf, lty="dashed", lwd=1)


####### Courbe EDF de la fonction G (Figure 10)

x11(7,3.5)
par(mfrow=c(1,2), cex.lab=0.9,mar=c(5,5,1,1))
plot(dis,GG, ty="l", xlab="y (mètres)",ylab="G(y)", lwd=2)
lines(dis, mg, lty="dashed")
lines(dis, Mg, lty="dashed")
lines(dis, Ftheo)

plot(Ftheo, GG, ty="l", lwd=2, xlab=expression(paste(G[th], "(y)")),
     ylab=expression(paste(hat(G), "(y)")))
lines(c(0,1),c(0,1), lwd=1)
lines(Ftheo, mg, lty="dashed", lwd=1)
lines(Ftheo, Mg, lty="dashed", lwd=1)





#####################################################
##                                                 ##
##                                                 ##
##             Fonction K(t) de Ripley             ##
##                                                 ##
##                                                 ##
#####################################################


### Calcul des fonctions K(t) et L(t) (Figure 12)
dis <- seq(1,10000,length=100)
```

```
Kobs <- khat(fac,pol, dis)
k <- list()
for (i in 1:99) {
  spts <- csr(pol, np)
  k[[i]] <- khat(spts, pol, dis)
}
k <- do.call("rbind", k)
mk <- apply(k, 2, min)
Mk <- apply(k, 2, max)
Ktheo <- pi*(dis^2)

LL <- sqrt(Kobs/pi) - dis
ml <- sqrt(mk/pi) - dis
Ml <- sqrt(Mk/pi) - dis


### Graphique des fonctions K(t) et L(t) (Figure 12)
x11(7,3.5)
par(mfrow=c(1,2), cex.lab=0.9,mar=c(5,5,1,1))
plot(dis,Kobs, ty="l", xlab="t (mètres)",ylab="K(t)", lwd=2)
lines(dis, mk, lty="dashed")
lines(dis, Mk, lty="dashed")
lines(dis, Ktheo)

plot(dis,LL, ty="l", xlab="t (mètres)",
     ylab=expression(L(t) == sqrt(K(t)/pi) - t), lwd=2)
lines(dis, ml, lty="dashed")
lines(dis, Ml, lty="dashed")
abline(h=0)




######################################################
##                                                  ##
##                                                  ##
##     Simulation de trois processus de points      ##
##                                                  ##
##                                                  ##
######################################################


par(mfrow=c(1,3), mar=c(0.1,0.1,0.1,0.1))


####### Le processus de Poisson (Figure 13A)
pol <- matrix(c(0,1,1,0,0,0,1,1),ncol=2)
set.seed(800)
sim1 <- csr(pol,500)
plot(sim1, pch=21, asp=1, bg="white")



####### Le processus de Poisson inhomogène (Figure 13B)

## Création de la surface de tendances:
## Simulation d'un gradient
uu <- ascgen(sim1, nrcol=500)
foo <- function(x)
  1-x
xyc <- getXYcoords(uu)
```

```
ii <- apply(uu,1, function(x) 1-xyc$x)
ii <- getascattr(uu,ii)

## On pose lambda max = 1
set.seed(716)
sim2<-csr(pol,1000)
jo <- join.asc(sim2,ii)
sim2 <- sim2[!is.na(jo),]
jo <- jo[!is.na(jo)]
cons <- runif(length(jo))
sim2 <- sim2[cons<=jo,]
image(ii)
points(sim2, pch=21, asp=1, bg="white")


####### Le processus de Neyman-Scott (Figure 13C)

sim3 <- pcp.sim(50, 10, 0.001, pol)
x11(9,3)
par(mfrow=c(1,3), mar=c(0.1,0.1,0.1,0.1))
plot(sim1, pch=21, asp=1, bg="white", axes=FALSE, cex=1.5)
box()
image(ii, axes=F, col=gray((240:50)/256))
box()
points(sim2, pch=21, asp=1, bg="white", cex=1.5)
plot(sim3, pch=21, asp=1, bg="white", axes=FALSE, cex=1.5)
box()
```

Annexe 11

Commandes du logiciel R utilisées pour construire les figures du chapitre 6

Commandes du logiciel R utilisées pour construire les figures du chapitre 6

Nous présentons ci-dessous les commandes que nous avons utilisées pour construire les figures du chapitre 6 avec le logiciel R. Les commentaires sont précédés d'un #.

Notons que le jeu de données utilisé dans le chapitre 6 n'est pas disponible dans **adehabitat**. En revanche, un échantillon de ces données est présent dans le jeu de données `chamois` de cette bibliothèque. Nous reproduisons donc les commandes utilisées dans le chapitre 6 sur cet échantillon.

```
######################################################
##                                                  ##
##                                                  ##
##            Chargement des données                ##
##                                                  ##
##                                                  ##
######################################################


library(adehabitat)
data(chamois)
a <- chamois$locs
kasc <- chamois$map

## diminution de la résolution de la carte
kasc <- lowres(kasc)

## conversion de la végétation en indicatrices de classes
kascb <- convnum(kasc)
map <- kascb$kasc

## Suppression de la forêt
map[[1]] <- NULL

## On ne conserve que les localisations tombant dans la
## zone d'étude
fac <- a[!is.na(join.asc(a,getkasc(map,2))),]



######################################################
##                                                  ##
##                                                  ##
##                   Les cartes                     ##
##                                                  ##
##                                                  ##
######################################################

## Les cartes des variables (Figure 18A)
image(map)
```

```
## La carte ponctuelle (Figure 18B)
cont <- getcontour(getkasc(map,1))
plot(cont, colpol="grey")
points(fac, pch=16)



####################################################
##                                                ##
##                                                ##
##       Exploration préliminaire des données       ##
##                                                ##
##                                                ##
####################################################


## ACP des pixels disponibles
lidf <- kasc2df(map)
df <- lidf$tab
pc1 <- dudi.pca(df,scan=F)
cartefac <- df2kasc(pc1$li, lidf$index, map)


#### Résultats de l'ACP (Figure 19)

## Diagramme en bâtons des valeurs propres
barplot(pc1$eig)

## Représentation des scores des pixels
opar <- par(mar=c(0,0,0,0))
s.label(pc1$li, clab=0, cpo=0,xlim=c(-5,4),
        ylim=c(-4,5))
points(pc1$li, col="grey", pch=16)
## Dénombrement des localisations dans chaque pixel
cpo <- count.points(as.data.frame(fac), map)
cpob <- c(cpo)[lidf$index]
ls <- pc1$li[cpob>0,]
points(ls, col="black", pch=16, cex=1.3)
abline(h=0,v=0)
par(opar)

## Cercle des corrélation
s.corcircle(pc1$co)

## Représentation de la niche
pcav <- mcp(pc1$li, id=rep(1,nrow(pc1$li)))
pcus <- mcp(ls, id=rep(1,nrow(ls)))
par(mar=c(0,0,0,0))
s.label(pc1$li, clab=0, cpo=0,xlim=c(-5,4), ylim=c(-4,5))
polygon(pcav[,2:3], col=grey(0.8))
polygon(pcus[,2:3], col=grey(0.1))

## histogrammes de la niche
histniche(map, fac)
```

```
######################################################
##                                                  ##
##                                                  ##
##                      ENFA                        ##
##                                                  ##
##                                                  ##
######################################################


## Résultats de l'ENFA (Figure 22)
dataenfa1 <- data2enfa(map, fac)
enfa1 <- enfa(dataenfa1$tab, dataenfa1$pr, scan=F,nf=3)
barplot(enfa1$s)
hist(enfa1)
scatter(enfa1, pts=T)
scatter(enfa1)



######################################################
##                                                  ##
##                                                  ##
##          Cartes de qualité de l'habitat          ##
##                                                  ##
##                                                  ##
######################################################


## Programmation d'une fonction pour appliquer
## l'algorithme BIOCLIM (non disponible dans adehabitat)

bioclim <- function(kasc, pts, nquantil=100)
  {
    x <- pts
    if (ncol(x)!=2)
      stop("x should have two columns")
    if (!inherits(kasc,"kasc"))
      stop("kasc should be of class \"kasc\"")
    if (any(apply(kasc,2,is.factor)))
      stop("not implemented for factors")
    df <- join.kasc(x,kasc)
    dfav <- kasc2df(kasc)
    foo <- function(i)
      {
        us <- df[,i]
        av <- dfav$tab[,i]
        p <- seq(0,1,length=nquantil)
        u <- quantile(us, p)
        p <- unlist(lapply(p, function(x) {
          if (x>0.5) return(1-x)
          return(x)}))
        p <- 2*p
        ind <- 1:length(u)
        lo <- unlist(lapply(lapply(av, function(x) u>x),
                            function(x) {
                              tutu <- ind[x]
                              if (length(tutu>0))
                                toto <- p[min(ind[x])] else return(0)
                              return(toto)
                            }))
        up <- unlist(lapply(lapply(av, function(x) u<x),
```

```
                          function(x) {
                            tutu <- ind[x]
                            if (length(tutu>0))
                              toto <- p[max(ind[x])] else return(0)
                            return(toto)
                          }))
        me <- median(us)
        res <- unlist(lapply(1:length(av), function(x) {
          if (av[x]<=me) return(lo[x])
          if (av[x] >me) return(up[x])
        }))
        return(res)
      }
    res <- lapply(1:ncol(dfav$tab),foo)
    o <- apply(do.call("cbind", res),1,min)

    o <- data.frame(o,o)
    o <- df2kasc(o,dfav$index,kasc)
    res <- getkasc(o,1)
    return(res)
  }



## Programmation d'une fonction pour construire
## la carte des moyennes géométriques
## calculées sur axes de l'ENFA (non disponible
## dans adehabitat)

cartoENFA1 <- function(en, daenf)
  {
    za <- as.matrix(en$li)
    pr <- en$pr/sum(en$pr)
    zu <- za*sqrt(pr)
    poids <- diag(1/((t(zu)%*%zu )/(t(za)%*%za/nrow(za))))
    poids=poids/sum(poids)
    zu <- za[pr>0,]
    res <- lapply(1:nrow(za), function(i) {
      unlist(lapply(1:nrow(zu), function(j) {
        sum(poids*((za[i,]-zu[j,])^2))
      }))})

    resb <- unlist(lapply(res, function(x) {
      pui <- 1/length(x)
      return(prod(x)^pui)}))

    resc <- data.frame(resb, resb)
    so <- getkasc(df2kasc(resc, daenf$index, daenf$attr),1)
    return(so)
  }


## Estimation des cartes

## BIOCLIM
bioc <- bioclim(map,fac)

## DOMAIN
dom <- 1-domain(map, fac)

## Moyenne géométrique sur axes de l'ENFA
```

```
cENFA <- cartoENFA1(enfa1, dataenfa1)

## Distances de Mahalanobis
maha <- mahasuhab(map,fac)

## Distances de Mahalanobis sur axes de l'ENFA
maha2 <- predict(enfa1, dataenfa1$index, dataenfa1$attr)


## création du kasc pour les cartes finales (Figure 23)
hasu <- as.kasc(list(BIOCLIM = 1-bioc, DOMAIN=dom,
                     "Moyenne géométrique sur axes de l'ENFA" = cENFA,
                     Mahalanobis=maha,
                     "Mahalanobis sur axes de l\'ENFA" = maha2))
image(hasu)
```

<u>Annexe 12</u>

Article 9 :

“The package adehabitat for the R software: a tool for the analysis of space and habitat use by animals”

Clément Calenge.

Ecological Modelling, soumis.

The package adehabitat for the R software: a tool for the analysis of space and habitat use by animals

Clément Calenge [a,1]

[a] *UMR CNRS 5558, Laboratoire de Biométrie et de Biologie Evolutive, Université Claude Bernard Lyon 1, 69622 Villeurbanne Cedex, France.*

---

[1] Corresponding author. Tel.: +33-04-72-43-27-56; fax.: +33-04-72-43-13-88.

E-mail address: calenge@biomserv.univ-lyon1.fr (C. Calenge)

Abstract

The practical analysis of space use and habitat selection by animals is often a problem due to the lack of well-designed programs. We present here the "adehabitat" package for the R software, which offers basic GIS (Geographic Information System) functions, methods to analyse radio-tracking data and habitat selection by wildlife, and interfaces with other R packages. Both the software and the package can be downloaded freely on the internet. Because the functions of this package can be combined with other functions of R, adehabitat provides a powerful environment for the analysis of the space and habitat use.

The study of the relationships between animals and their environment is one of the main issues in Ecology. It implies the analysis of the use of space by animals, of the interactions between animals, and of the relationships between animals and their habitat. The recent development of Geographic Information Systems (GIS) has made easier this type of study, and especially the study of habitat selection, by taking into account more explicitly the spatial dimension of the data in the analyses (Manly et al., 2002). This expanded the analysis possibilities, and as a result numerous methods have been developed during the last decade in this field of Ecology (e.g. Guisan and Zimmermann, 2000).

But the practical data analysis remains a problem. First, the analyst has to perform spatial analyses within a GIS (computation of buffers, estimation of home ranges, etc.), and second, he needs a statistical software to analyze the data more deeply. The interface between GIS and statistical software is often missing and can make the analysis long and heavy. In addition, the use of commercial statistical program in Ecology science has been criticized by several authors (Chessel, 1992; Tufto and Cavallini, 2005). Indeed, such “canned” programs are not open enough. They provide only standard statistical procedures, and limit the development of new approaches.

In this framework, open-source software may be viewed as an alternative (Tufto and Cavallini, 2005). Such programs are distributed freely under a licence known as the GNU General Public Licence (GPL; Free Software Foundation, Inc., 1991), which also relies on standard copyright laws. All the source code is available, so that a user can make changes inside the program. The modified version is also free and available. This usually results in very efficient programs, because they are built and modified by their users.

Tufto and Cavallini still noted that the awareness of this phenomenon was weaker among biologists than researchers belonging to other fields of science, despite its undeniable desirable qualities. They recommended the use of free software, and listed the programs that may be of great interest, giving a special mention to the R software (R Development Core Team, 2005).

Our propose is to present, after a brief description of the R software, the package we developed to analyze the use of space and habitat selection by animals: adehabitat. It gatheres a set of methods, through R functions, intended to provide tools for biologists.

1. The R software

Ross Ihaka and Robert Gentleman, from the Auckland University, developed the R software to provide a statistical environment to their laboratory in 1992, and based this software on the S language. This one was , invented at the AT&T Bell Laboratories by John Chambers and his colleagues during the mid-70's. They wanted to encourage the user to "slide into programming, perhaps without noticing" (Chambers, 1998), and to ensure that a user can easily implement his own tecniques.

With the help of Martin Mächler (working at the ETH Zürich), Ihaka and Gentleman released this software as free open-source software in 1995 (Ihaka and Gentleman 1996). Presently, the "Comprehensive R Archive Network" (CRAN) is the core of an increasingly growing R community (Hornik and Leisch, 2001); it is the central repository for material related to R, including packages of functions (518 at the time of writing), tutorials and discussion forums (URL: http://cran.r-project.org/). The use of R implies the learning of a command language. At first sight, this aspect may seem to be a drawback, but it has actually numerous desirable qualities. First, this language is very intuitive (e.g. a t-test is carried out with the function `t.test()`, a general linear model is carried out using the function `glm()`, etc.). Numerous tutorials are available on the internet to learn quickly this language.

Several packages are available for the analysis of ecological data. Among them, the "ade4" package allows the analysis of ecological and environmental data in the framework of

Euclidean Exploratory methods (Chessel et al., 2004), the "vegan" package is intended for the vegetation and community analysis, and the "GRASP" package aims at building predictive models of the distribution range of species in biogeography, using generalized regression and spatial prediction (Lehmann et al., 2002).

2. The adehabitat package

The adehabitat package has been developed to study the space use and habitat selection by wildlife. It contains about 100 functions giving tools frequently used in this field of research. The graphical possibilities of this package (Fig. 1), and the combination of the adehabitat functions with the powerful analysis environment provided by R allow the users to design a large diversity of analyses of the relationships between animals and their environment. A quick learning can be carried out with the tutorial included in the package.

The adehabitat package includes functions that create an interface with GIS commonly used in Ecology, such as Arcview GIS (ESRI, 1996) or the free program Landserf (available at www.soi.city.ac.uk/~jwo/landserf/). Single maps can be imported within R, and then combined into multilayer maps; these maps are grids of pixels on which several variables are measured, allowing multivariate analysis of habitat use by animals. Several basic GIS functions are also available in adehabitat. It is for example possible to compute buffers around lines or points, to determine the habitat composition at given points, to select sub-area on a map, to define masks, to reduce the resolution of a map, and so on. In addition, several functions provide an interface with the spatial classes of the sp package which is itself an interface toward other R packages dealing with spatial data (Pebesma and Bivand, 2005), another efficient way to analyze data.

Common analyses of habitat selection may also be carried out using adehabitat. Thus, several functions allow the computation of selection ratios (Manly et al., 2002), compositional analysis (Aebischer et al., 1993), K-select analysis (Calenge et al., 2005) or Ecological Niche Factor Analysis (ENFA, Hirzel et al., 2002). Moreover, habitat suitability maps can be computed using Mahalanobis distances (Clark et al., 1993), the DOMAIN algorithm (Skov, 2000), or the ENFA. Resource selection functions may also be fitted, using the spatial classes of adehabitat and the powerful modelling capabilities of R. To take into account the multivariate aspect of habitat (Hall et al., 1997), this package also provides an interface with the package ade4 (Chessel et al., 2004).

Adehabitat also includes several tools suitable for the analysis of radio-tracking data. Home range estimators include the minimum convex polygon (Mohr, 1947) the kernel estimator (Worton, 1989), the nearest neighbor convex hull (Getz and Wilmers, 2004), or the grid method (Siniff and Tester, 1965). And so, the estimates can be used in the analyses of spatial interactions between animals, habitat selection, etc. Schoener's ratio can be computed to detect temporal autocorrelation in the relocations (Swihart and Slade, 1985), and basic statistics to analyze animal movements such as speed or turning angles (Turchin, 1998).

3. Discussion

The main objective of the adehabitat package is to provide tools for analyzing the relationships between animals and their environment. Because the R community is very active, the package can be improved frequently, adding new available functions. Thus, since its first release in September 2004, two additional versions of the package have been submitted on CRAN as a result of the suggestions made by the users concerning the methods that would find their place in adehabitat. For instance, the recent development of GPS (Global Positioning System) to monitor space use by animals has raised a number of questions among ecologists (autocorrelation, how to analyze such trajectories, etc). In view of this situation, Paolo Cavallini (working at Faunalia, Italy) created a discussion forum to make the communication easier between scientists of various fields interested by the analysis of animal movement (URL: www.faunalia.com/animov/). Several functions of the package have been added or modified following discussions that arose on this forum, e.g. the kernel Brownian bridge estimator of the home range (Bullard, 1991) or the nearest neighbour convex hull estimator of the home range (Getz and Wilmers, 2004).

The functions of adehabitat, and more generally, the R functions are well-documented on the help pages. These pages describe the arguments and options needed by the functions, give examples of use, and sometimes the context in which they were developed and the context in which they are to be used. Thus, R is more than a simple technical tool, and becomes integral part of the development of methods. This illustrates well the philosophy behind the S language, which is, owing to the nice expression of Chambers (1998), “to turn ideas into software”. R is therefore the biometrical tool par excellence. As the interdisciplinarity emerges today as an important aspect of science (e.g. Morillo et al., 2004), the R software and the adehabitat package will be of major use to wildlife biometricians who are concerned by the use of space and habitat selection by animals.

Acknowledgments

We would like to thank the “Office national de la chasse et de la faune sauvage” for their financial support. We warmly thank Anne-Béatrice Dufour (University of Lyon, France) for her unvaluable comments on earlier drafts on this manuscript. We are also grateful to Mathieu Basille who programmed the ENFA and related functions for adehabitat, and suggested many improvements to bring to the package; to Daniel Chessel for his helpful advices about “good programming practice” and all the users who send us their comments and contributed to the improvement to the package.

References

Aebischer, N.J., Robertson, P.A., Kenward, R.E., 1993. Compositional analysis of habitat use from animal radio-tracking data. Ecology 74, 1313-1325.

Bullard, F., 1991. Estimating the home range of an animal: a Brownian bridge approach. Msc. Thesis. Johns Hopkins University, Chapel Hill, North Carolina.

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

Chambers, J.M., 1998. Programming with data. Springer, New York.

Chessel, D., 1992. Echanges interdisciplinaires en analyse de données écologiques. Professoral thesis. Université Lyon 1, Villeurbanne, France.

Chessel, D., Dufour, A.B., Thioulouse, J., 2004. The ade4 package - I : One-table methods. R news 4, 5-10.

Clark, J. D., Dunn, J.E., Smith, K.G., 1993. A multivariate model of female black bear habitat use for a geographic information system. J. Wildl. Manage. 57, 519-526.

ESRI, 1996. Using Arcview GIS. The geographic information system for everyone. Environmental Systems Research Institute, Inc, USA.

Free Software Foundation, Inc., 1991. GNU General Public License, version 2. Available at: www.gnu.org/copy left/gpl.html

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecological modelling 135, 147-186.

Getz, W.M., Wilmers, C.C., 2004. A local nearest-neighbor convex-hull construction of home ranges and utilization distributions. Ecography 27, 489-505

Hall, L.S., Krausman, P.R., Morrison, M.L., 1997. The habitat concept and a plea for standard terminology. Wildl. Soc. Bull. 25, 173-182.

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

Hornik, K., Leisch., F., 2001. Vienna and R: Love, Marriage and the Future. Pages 61-70 *in* R. Dutter, editor. Festschrift. 50 Jahre Österreichische Statistische Gesellschaft. Österreichische Statistische Gesellschaft, Vienna, Autriche.

Ihaka, R., Gentleman, R., 1996. R: A Language for Data Analysis and Graphics. J. Comput. Graph. Stat. 5, 299-314.

Lehmann, A., Leathwick, J.R., Overton, J.M., 2002. GRASP v.2.2 User's Manual. Landcare Research, Hamilton.

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

Mohr, C.O., 1947. Table of equivalent populations of North American small mammals. Amer. Midl. Nat. 37, 223-249.

Morillo, F., Bordons, M., Gomez, I., 2003. Interdisciplinarity in Science: a tentative typology of disciplines and research areas. Journal of the American Society for Information Science and Technology 54, 1237-1249.

Pebesma, E.J., Bivand, R.S. 2005. S classes and Methods for Spatial data: the sp package. Unpublished report.

R Development Core Team, 2005. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. www.R-project.org

Siniff, D.B., Tester, T.W., 1965. Computer analysis of animal-movement data obtained by telemetry. Bioscience 15, 104-108.

Skov, F., 2000. Potential plant distribution mapping based on climatic similarity. Taxon 49, 503-515.

Swihart, R.K., Slade, N.A., 1985. Testing for independence of observations in animal movements. Ecology 66, 1176-1184.

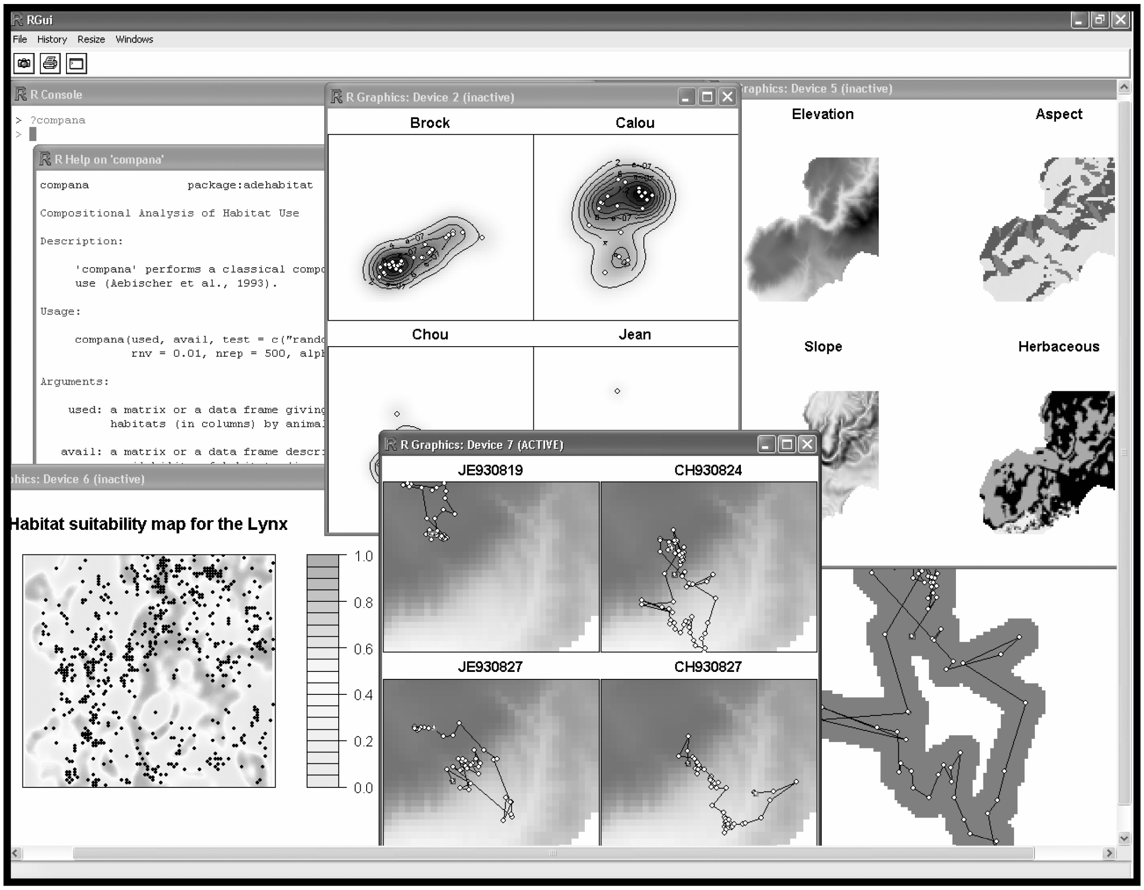
Tufto, J., Cavallini, P., 2005. Should wildlife biologists use free software ? Wildl. Biol. 11, 67-76.

Turchin, P., 1998. Quantitative analysis of movement, Sunderland, MA.

Worton, B.J., 1989. Kernel methods for estimating the utilization distribution in home-range studies. Ecology 70, 164-168.

Figures legend

Fig. 1. Capture screen of one session with R and the package adehabitat.



```
> ?compana
> 
```

```
compana              package:adehabitat

Compositional Analysis of Habitat Use

Description:

     'compana' performs a classical compo
     use (Aebischer et al., 1993).

Usage:

     compana(used, avail, test = c("rando
             rnv = 0.01, nrep = 500, alph

Arguments:

    used: a matrix or a data frame giving
          habitats (in columns) by animal

   avail: a matrix or a data frame descri
```

1

Annexe 13

Didacticiel 1 :

“Classes of map objects available in adehabitat”

Clément Calenge.

Didacticiel de la bibliothèque de fonctions **adehabitat**

Classes of map objects available in adehabitat

Clément Calenge

May 10, 2005

Contents

Chapter 1

Package objective

This package was primarily developed for the analysis of radio-tracking data. Common analyses are available for home-range estimation or study of habitat selection. However, **adehabitat** has been useful in other fields, such as biogeography (Spichiger *et al.*, 2004, The geographical zonation in the Neotropics of tree species characteristic of the Paraguay-Paraná Basin, *Journal of Biogeography*, 31, 1489-1501). Basically, this package was developed to provide an interface between Geographic Information Systems (G.I.S.) and the R package **ade4**, useful for the multivariate analysis of ecological data (Chessel *et al.*, 2004, The ade4 package - I : One table methods, *R News*, 4/1, 5-10).

Roughly speaking, the functions available in this package are general enough to provide useful tools to ecologists who want to match one or several sets of points (animals, plant species, etc.) with a set of raster maps. Some basic GIS operations are available (labeling of connected features, spatial join of a set of points with a set of maps, etc.). In this document, we present the different classes of objects used in **adehabitat** to store maps.

First of all, you need to load the package:

```
> library(adehabitat)
```

Chapter 2

The storage mode of raster maps in adehabitat

2.1 Single maps : the class `"asc"`

Two kinds of maps must be distinguished in geographical analysis (see Figure 2.1):

- *the vector maps* divide the area into a set of polygons of various sizes, with different attributes;
- *the raster maps* divide the area of interest into a set of square cells, the pixels, all of the same size, each pixel having a value for the mapped variable.

Figure 2.1: (left) a vector map; (right) a raster map.

Only raster maps can be imported in R with **adehabitat**. When vector maps are to be imported into R , you may consider the wonderful package **maptools**. Wildlife scientists often prefer to work with raster maps, because several environmental variables are often under study. With raster maps, all the pixels have the same shape and the same area. This equal importance of the sampling units makes easier the analyses of such multivariate data.

Raster maps may be either of type `numeric` (a numerical variable is mapped, *e.g.* the elevation) or of type `factor` (a categorical variable is mapped, *e.g.* the vegetation type). Both types of maps are handled with **adehabitat**.

These maps are stored in R as matrices of class `"asc"`. Each cell of the matrix corresponds to a pixel of the map. These maps have attributes related to the position and the resolution of the map (see Figure 2.2):

- `xll` : the x coordinate of the centre of the lower left pixel of the map;
- `yll` : the y coordinate of the centre of the lower left pixel of the map;
- `cellsize` : the size of the side of a pixel on the studied map.

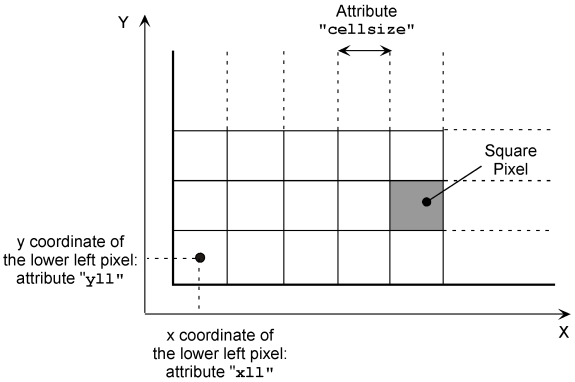


Figure 2.2: Detail of a raster map of class `"asc"` illustrating the attributes related to its position and its resolution.

Two other attributes are related to the mapped variables

- `type` : either `"numeric"` or `"factor"`;
- `levels` : if the type of the map is `"factor"`, the levels of the factor are considered.

Objects of class `"asc"` can be created using the function `as.asc`. We generate here a map with random numbers (Figure 2.3):

```
> mat <- matrix(rnorm(10000), 100, 100)
> asc <- as.asc(mat)
> image(asc)
> box()
```

However, the most common way to create objects of class `"asc"` is to import Arcview ASCII raster files (files with the extension ".asc"), with the function `import.asc()`. For example, **adehabitat** contains a raster map named "elevation.asc" in the directory "ascfiles". The path for this file can be obtained with the command:

```
> (path.to.file <- paste(system.file(package = "adehabitat"), "ascfiles/elevation.asc",
+     sep = "/"))
```

```
[1] "C:/Rdev/R-2.1.0/library/adehabitat/ascfiles/elevation.asc"
```

Of course, this path may vary from one machine to another. The map is then imported into R (Figure 2.4):

```
> el <- import.asc(path.to.file)
> image(el, main = "Elevation")
```

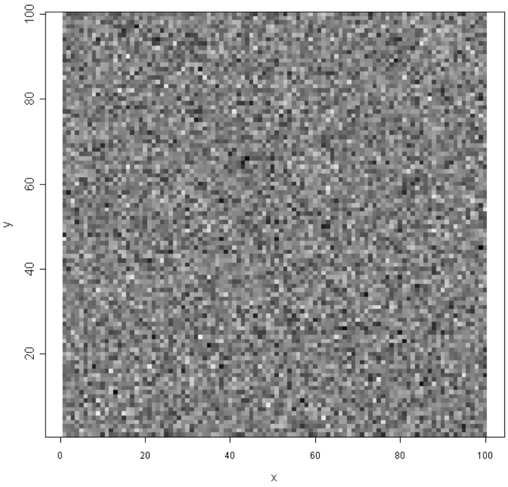


Figure 2.3: An object of class `"asc"` generated with `as.asc()` on a matrix of random numbers.

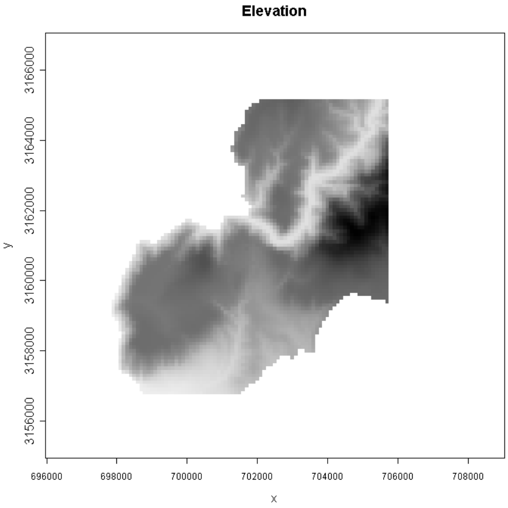


Figure 2.4: A raster map of type `"numeric"` imported into R with the function `import.asc()`(Map of the elevation at Puéchabon, South of France).

Conversely, the exportation of a matrix of class `"asc"` can be done using the function `export.asc()`:

```
> export.asc(el, "toto.asc")
```

For factor maps, it is also necessary to define the levels of the variable. For example, consider the map “aspect.asc”, located in the directory “ascfiles” of the package. This map describes the aspect of the study area, and has four levels (`North`, `East`, `West`, `South`). However, these levels are not stored in the ASCII raster file:

```
> (path.to.map <- paste(system.file(package = "adehabitat"), "ascfiles/aspect.asc",
+     sep = "/"))

[1] "C:/Rdev/R-2.1.0/library/adehabitat/ascfiles/aspect.asc"
```

```
> asp <- import.asc(path.to.map, type = "factor")
> levels(asp)

[1] "1" "2" "3" "4"
```

It is therefore necessary to specify the labels associated with each level of the variable when importing a factor map:

```
> asp <- import.asc(path.to.map, type = "factor", lev = c("North",
+     "East", "West", "South"))
> levels(asp)

[1] "North" "East"  "West"  "South"
```

Another way of specifying the levels of the factor map is to export the theme table from Arcview. To proceed, select the theme in Arcview, then choose the menu Theme -> Table, and finally export the table as a delimited text file from the menu File -> Export. The file “aspect.txt” is located in the directory “ascfile” of the package:

```
> (path.to.table <- paste(system.file(package = "adehabitat"),
+     "ascfiles/aspect.txt", sep = "/"))

[1] "C:/Rdev/R-2.1.0/library/adehabitat/ascfiles/aspect.txt"
```

```
> file.show(path.to.table)
```

This text file has three columns, which are separated by commas.

```
"Value","Count","NewField1"
 1,537,North
 2,1504,East
 3,1262,South
 4,1076,West
```

The first column gives the levels of the map, and the third column gives the corresponding labels (The second column is no interest for our purpose). This file may then be specified when importing the map:

```
> asp <- import.asc(path.to.map, type = "factor", lev = path.to.table,
+     levnb = 1, labnb = 3)
> levels(asp)

[1] "North" "East"  "South" "West"
```

```
> co <- colasc(asp, North = "blue", East = "yellow", West = "orange",
+     South = "red")
```

`colasc()` links a vector of colors with each level of `asp`. Then, the map is displayed (Figure 2.5):

```
> image(asp, clfac = co)
> legend(696662, 3166028, legend = levels(asp), fill = co)
```

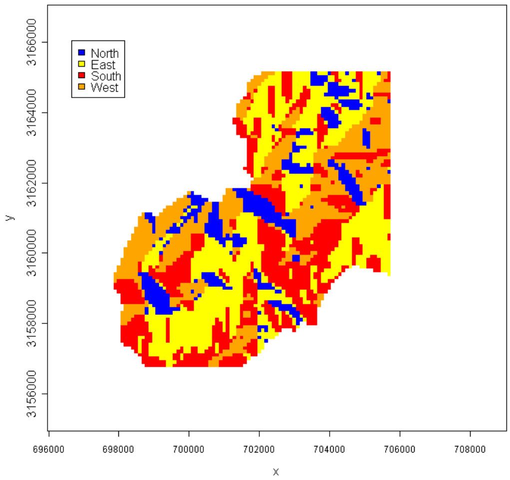


Figure 2.5: A raster map of type `"factor"` imported into R with the function `import.asc()`(map of the aspect at Puéchabon, South of France).

Finally, there are other ways to create objects of class `"asc"`: the functions `asc2im()` and `im2asc()` provide an interface with the class `"im"` of the package **spatstat**.The functions `spixdf2kasc()`, `kasc2spixdf()`, and `asc2spixdf()` provide an interface with spatial classes of the package **sp**.

Note: The central place of this class of objects in **adehabitat** arises from the large availability of **Arcview GIS** among ecologists (`http://www.esri.com/`). The class `"asc"` originates from the ASCII raster maps created by Arcview GIS (under the File menu: Export Data source). Note, however, that the format coming from other GIS softwares may be converted to ASCII raster maps using the free software **Landserf**, downloadable at the following URL: `http://www.soi.city.ac.uk/~jwo/landserf/`.

2.2 Multiple raster maps in a unique object: the class `"kasc"`

Ecologists are usually interested in several environmental variables. Therefore, they have to deal with a collection of K raster maps of the same area. These maps often have the same attributes (same number of rows and of columns, same resolution, same coordinates for the lower left pixel). Such a collection of maps may be stored in objects of class `"kasc"`.

This class inherits from the class `"data.frame"`. Each column of this data frame corresponds to the map of one variable, with one row per pixel of the map (Figure 2.6).

Thus, all sub-setting rules that are applied to the class `"data.frame"` can also be applied to the class `"kasc"`. Objects of class `"kasc"` have the same attributes as objects of the class `"asc"` (described in the previous section), but they also have two additional attributes:

- `nrow`: the number of rows of the raster map. *Warning!* the number of rows of the map is equal to the number of columns of the matrix of class `"asc"` coding this map;

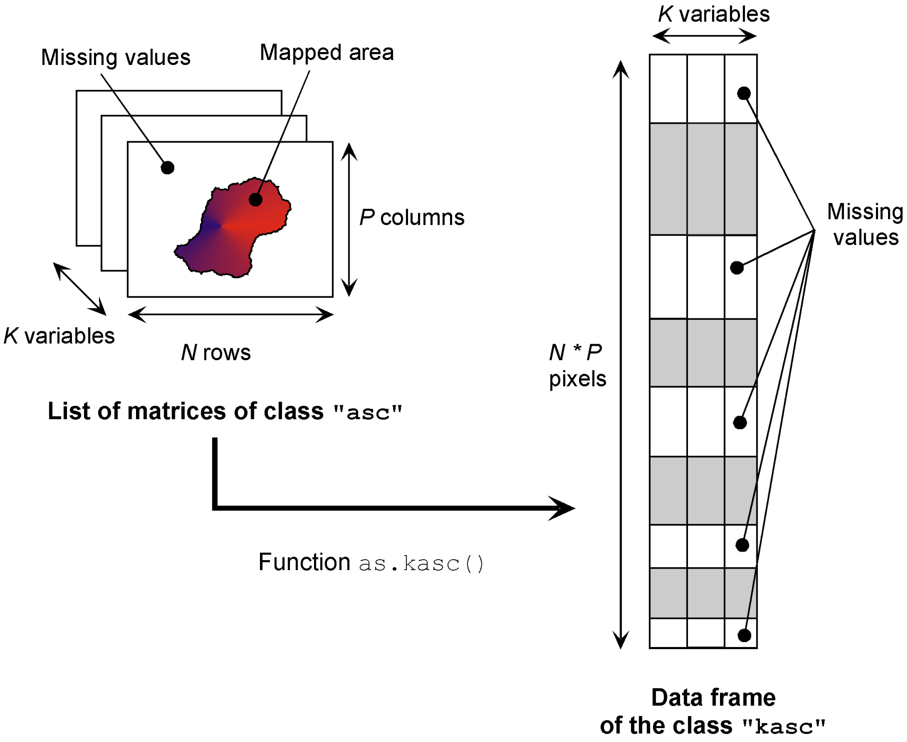


Figure 2.6: A data frame of the class `"kasc"` is created from several raster maps of class `"asc"`. Note that when the mapped area does not fit the whole map (which is the case here), this leads to a considerable amount of missing values in the object of class `kasc`.

- `ncol`: the number of columns of the raster map. *Warning!* the number of columns of the map is equal to the number of rows of the matrix of class `"asc"` coding this map.

An example of object of class `"kasc"` is provided in the data set `puechabon`. The component named `kasc` of this list is a map of class `"kasc"`:

```
> data(puechabon)
> kasc <- puechabon$kasc
> image(kasc)
```

The image returned by these instructions is displayed Figure 2.7.

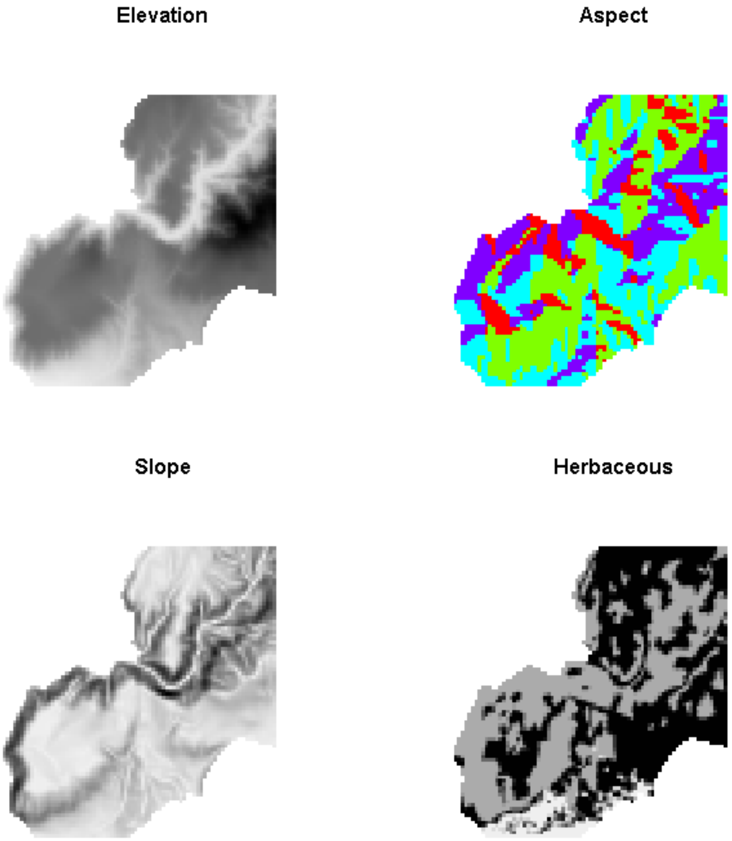


Figure 2.7: Example of object of class `"kasc"` (maps of Elevation, Slope, Aspect and Herbaceous cover at Puéchabon, South of France).

The most common way to create objects of class `"kasc"` is to use the function `as.kasc()`. In this example, don't forget that the maps `el` and `asp` have already been imported in the previous section. Note that each element of the list passed to the function `as.kasc` should be named:

```
> (obj <- as.kasc(list(Elevation = el, Aspect = asp)))
```

```
Raster map of class "kasc":
Cell size:  100
Number of rows:  121
Number of columns:  111

Variables measured:
1. Elevation: numeric
```

```
2. Aspect: factor
```

2.3 Basic operations with maps of classes "asc" and "kasc"

2.3.1 Introduction

We describe here some basic operations that can be done on maps with **adehabitat**. We mainly use the data set `puechabon` available in the package **adehabitat**. This data set stores information on the use of space by four wild boars (*Sus scrofa*) monitored by radio-tracking in the South of France (see the help page of this data set for further information and source). We use here the components: (i) `puechabon$kasc`, an object of class `"kasc"` that describes several variables on the study area; and (ii) `puechabon$locs`, a data frame containing the coordinates of the relocations of the wild boars resting sites in summer, as well as information on wild boars in factors `Name`, `Sex`, `Age`. We first load the data set:

```
> data(puechabon)
> puechabon$kasc
```

```
Raster map of class "kasc":
Cell size:  100
Number of rows:  121
Number of columns:  111

Variables measured:
1. Elevation: numeric
2. Aspect: factor
3. Slope: numeric
4. Herbaceous: numeric
```

```
> puechabon$locs[1:4, ]
```

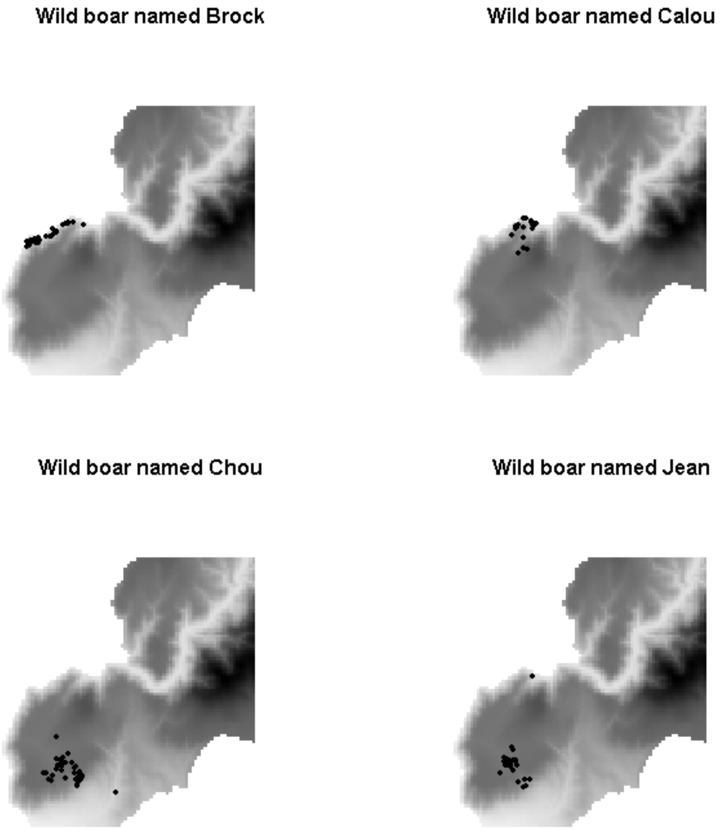
```
   Name Age Sex      X       Y   Date
1 Brock   2   1 699889 3161559 930701
2 Brock   2   1 700046 3161541 930703
3 Brock   2   1 698840 3161033 930706
4 Brock   2   1 699809 3161496 930707
```

A general view of the data is displayed using (Figure 2.8):

```
> el <- getkasc(puechabon$kasc, "Elevation")
> opar <- par(mfrow = c(2, 2), mar = c(0, 0, 4, 0))
> for (i in levels(puechabon$locs$Name)) {
+     image(el, main = paste("Wild boar named", i), axes = FALSE)
+     points(puechabon$locs[puechabon$locs$Name == i, c("X", "Y")],
+         pch = 16)
+ }
> par(opar)
```

In this section, we also use another data set, `chamois`. This data set stores locations of chamois (*Rupicapra rupicapra*) during the hunting seasons of two years (1992 and 1997), collected in the Chartreuse mountain (French Alps) by census operations (see the help page of this data set for further information and source). This list has two elements: (i) `chamois$map`, an object of class `"kasc"` that describes three habitat variables on the study area; and (ii) `chamois$locs`, a data frame containing the coordinates of the locations of the chamois. We also load this data set:

```
> data(chamois)
> chamois$map
```


Wild boar named Brock

Wild boar named Calou

Wild boar named Chou

Wild boar named Jean

Figure 2.8: Distribution of the relocations of four wild boars monitored using radio-tracking at Puéchabon, South of France).

```
Raster map of class "kasc":
Cell size:  50
Number of rows:  353
Number of columns:  353

Variables measured:
1. Vegetation: factor
2. Disteco: numeric
3. Slope: numeric
```

```
> chamois$locs[1:4, ]
```

```
           X       Y
72  862793.5 2038625
180 867550.2 2047080
17  872687.6 2046367
277 866479.5 2043730
```

A general view of the data is displayed using (Figures 2.9):

```
> sl <- getkasc(chamois$map, "Slope")
> image(sl, main = "Distribution of chamois occurrences in the Chartreuse mountain")
> points(chamois$locs, pch = 16)
```

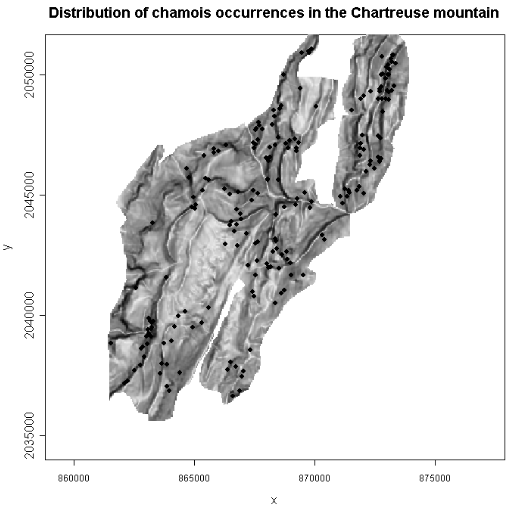


Figure 2.9: Distribution of Chamois occurrences collected during two years (1992 and 1997) by census operation in the Chartreuse mountain (French Alps).

2.3.2 Getting a component of an object of class `"kasc"`

Several functions in **adehabitat** return objects of class `"kasc"`. For example, the user can convert one of the variables of the object of class `"kasc"` into a matrix of class `"asc"`, either for further analyses or for exportation toward a Geographic Information System. This is done with the help of the function `getkasc()`. For example, using the class object `"kasc"` of the data set `puechabon`:

```
> kasc <- puechabon$kasc
> (el <- getkasc(kasc, "Elevation"))
```

```
Raster map of class "asc":
Cell size:  100
Number of rows:  121
Number of columns:  111
Type:  numeric
```

`el` is the map of the elevation on the area. It is displayed in Figure 2.4.

2.3.3 Converting a map of type `"factor"` into several maps of distances

In some cases, It may be desired to compute map of distances to patches belonging to the different levels of a map of class `"asc"` and of type `"factor"`. For example, if the habitat selection by an animal is studied with a map of vegetation (with different levels: woods, open areas, crops, etc.), it may be desirable to take into account the fact that the animal select one habitat type (e.g. crops, for feeding) while staying close to another type (e.g. woods, for protection). In such cases, one may prefer to work with a set of map of distances to the habitat types, rather than with a single map of type `"factor"`.

The function `distfacmap()` performs this operation. For example, we use the map of the `"Aspect"` at Puéchabon (See Figure 2.5). One can compute the distance to the patches of each level (NorthEast, SouthEast, NorthWest, SouthWest; Figure 2.10):

```
> image(distfacmap(getkasc(puechabon$kasc, "Aspect")))
```

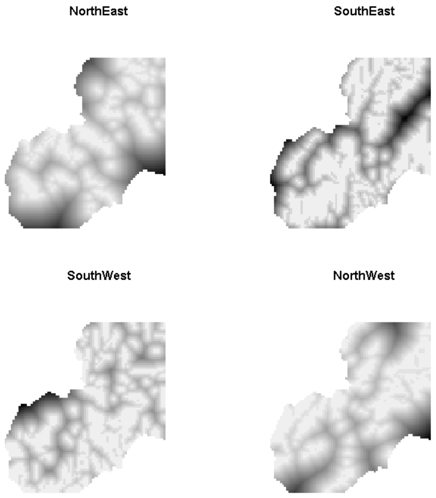


Figure 2.10: The distance to each level of the map of the Aspect at Puéchabon, South of France

2.3.4 Mathematical morphology

Some basic operations of mathematical morphology are available in **adehabitat**. Erosion and dilatation can be performed with the function `morphology()`. For example, using the map `el` created in the previous section:

```
> er8 <- morphology(el, operation = "erode", nt = 8)
> di8 <- morphology(el, operation = "dilate", nt = 8)
```

The maps `er8` and `di8` are objects of class `"asc"`, with pixels taking the value `1` if they belong to the mapped features, and missing values otherwise. The argument `nt` of the function indicates the number of times that the operation is to be processed (See Figure 2.11):

```
> image(di8, col = "black")
> image(el, col = "gray", add = TRUE)
> image(er8, col = "white", add = TRUE)
> arrows(704295, 3159355, 706588, 3157294, col = "red", lwd = 2,
+     code = 1)
> text(706240, 3156738, "Boundary of the study area")
```

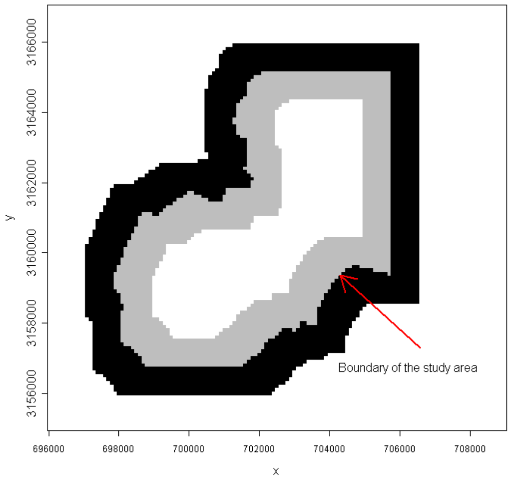


Figure 2.11: Example of morphological operations performed on the maps of the data set `puechabon` with the function `morphology()`. Buffer areas of 800 m have been defined inside and outside the study area.

2.3.5 Computing buffers around points

This operation, somewhat related to the previous one, can be processed using the function `buffer()`. For example, consider the component `locs` of the data set `puechabon`. This data frame contains the coordinates of the relocations of wild boars monitored using radio-tracking. Also consider the map `el`, created in the section 2.3.2:

```
> data(puechabon)
> puechabon$locs[1:4, ]
```

```
   Name Age Sex      X       Y   Date
1 Brock   2   1 699889 3161559 930701
2 Brock   2   1 700046 3161541 930703
3 Brock   2   1 698840 3161033 930706
4 Brock   2   1 699809 3161496 930707
```

A map of the relocations on the study area (Figure 2.12) is obtained with:

```
> image(el)
> points(puechabon$locs[, c("X", "Y")], pch = 16)
```

Here, the four wild boars are not distinguished. A buffer of 500 m is computed around each point (*e.g.* to take into account the lack of precision in the relocations, or to define areas that are available to the animals; Figure 2.13) :

```
> bu <- buffer(puechabon$locs[, c("X", "Y")], el, 500)
> image(bu)
> points(puechabon$locs[, c("X", "Y")], pch = 16)
```

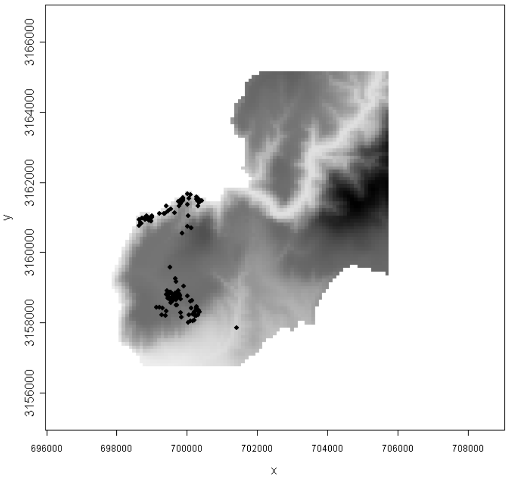


Figure 2.12: Relocations of 4 wild boars monitored using radio-tracking at Puéchabon (South of France).

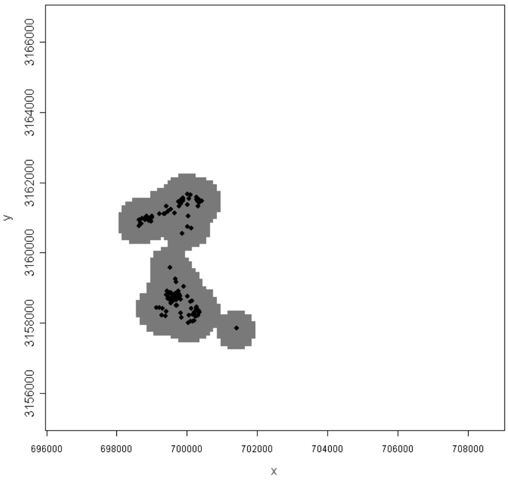


Figure 2.13: A buffer of 500 m around all relocations of the four boars monitored using radio-tracking. The buffer is computed using the function `buffer()`.

Note, that the map `el` could have been replaced by an object of class `"kasc"` to compute the same buffer (you may try to perform this operation by replacing `el` by `puechabon$kasc`). The map `bu` is an object of class `"asc"`, with pixels taking the value `1` if they belong to the mapped features, and missing values otherwise. It is easy after that to multiply the maps `el` and `bu` to measure the elevation in the neighbouring of the relocations (Figure 2.14). Thus:

```
> bubis <- bu * el
> mean(as.vector(bubis), na.rm = TRUE)

[1] 232.4520
```

```
> sd(as.vector(bubis), na.rm = TRUE)

[1] 69.06259
```

```
> image(bubis)
```

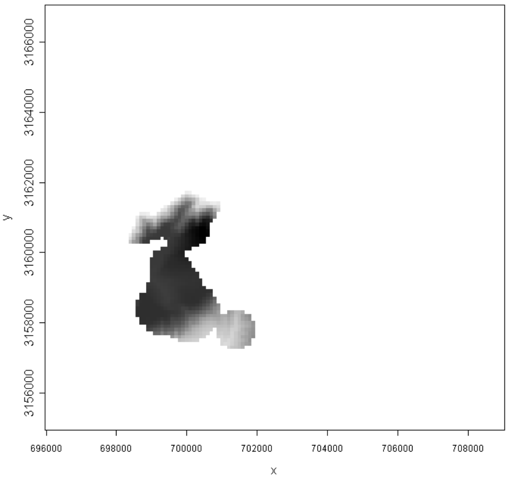


Figure 2.14: The elevation within 500 m from all relocations of the four wild boars monitored at Puéchabon (South of France) using radio-tracking.

Note that buffers can also be computed around lines, with the function `buffer.line()` (see the help page of this function).

2.3.6 Finding habitat features at every points

This operation can be processed using the function `join.asc()`. For example, using the map `el` and the component `locs` of the data set `puechabon`, it is possible to determine the elevation at each wild boar relocation in `puechabon$locs`:

```
> vec <- join.asc(puechabon$locs[, c("X", "Y")], el)
> length(vec)

[1] 119
```

```
> nrow(puechabon$locs)

[1] 119
```

```
> vec[1:10]

 [1] 108 172 121 113 101 127 102 108 104 151
```

`vec` contains the elevation of each relocation in `puechabon$locs`. It can further be used in statistical analyses, *e.g.* to study habitat selection by the boars. The function `join.kasc()` matchs one set of points with several maps (*i.e.* an object of class `kasc`):

```
> df <- join.kasc(puechabon$locs[, c("X", "Y")], puechabon$kasc)
> nrow(df)
[1] 119

> nrow(puechabon$locs)
[1] 119

> df[1:10, ]
   Elevation    Aspect    Slope Herbaceous
1        108 NorthWest 19.39472        0.2
2        172 NorthWest 23.67911        0.2
3        121 NorthEast 20.16249        0.2
4        113 NorthWest 22.39815        0.2
5        101 NorthWest 14.97099        0.2
6        127 NorthWest 16.82134        0.2
7        102 NorthEast 18.14621        0.2
8        108 NorthWest 18.08812        0.2
9        104 NorthEast 21.24609        0.2
10       151 NorthWest 29.24554        0.2
```

Each row of `df` gives the habitat composition for each relocation in `locs`.

2.3.7 Counting the number of points in each pixel of a map

This operation can be processed using `count.points()`. Consider again the map `el` and the wild boars relocations `puechabon$locs`:

```
> (cp <- count.points(puechabon$locs[, c("X", "Y")], el))
Raster map of class "asc":
Cell size:  100
Number of rows:  121
Number of columns:  111
Type:  numeric

> image(cp)
> box()
```

`cp` contains the number of relocations in each pixel of the map (Figure 2.15). It can be coerced into vector for a further analysis of the counts. When several sets of points are available (for example the locations of several animals monitored by radio-tracking, or several sets of species occurrences on the same area), the function `count.points.id()` counts, for each set, the number of points in each pixel of the map (Figure 2.16). For example, for the relocations of the wild boars `puechabon$locs`, the animals identity is indicated in the column `Name`:

```
> (cp <- count.points.id(puechabon$locs[, c("X", "Y")], puechabon$locs$Name,
+     el))
Raster map of class "kasc":
Cell size:  100
Number of rows:  121
Number of columns:  111

Variables measured:
1. Brock: numeric
2. Calou: numeric
3. Chou: numeric
4. Jean: numeric


> image(cp)
```

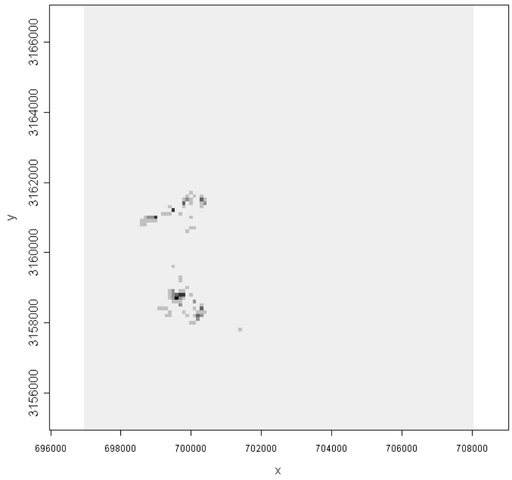


Figure 2.15: The number of wild boars relocations inside each pixel of the raster map `el` (numbered with the function `count.points()`).

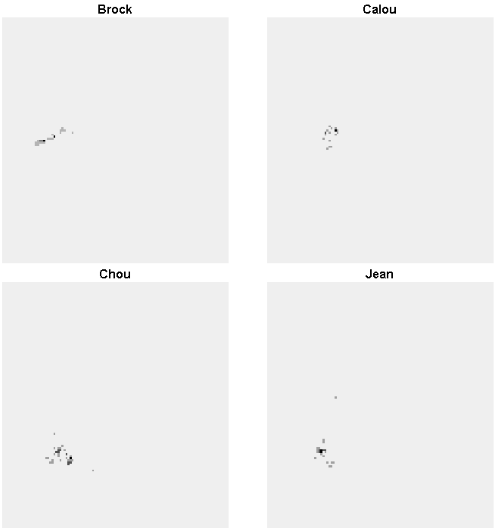


Figure 2.16: The number of relocations of each wild boar inside each pixel of the raster map `el` (computed with the function `count.points.id()`).

2.3.8 Counting the number of points in each cell of a virtual grid

This operation is very similar to the previous one, except that in this case, the raster map is not available to the analyst. Consider only the wild boars relocations `puechabon$locs`. A virtual grid can be superposed to the study area using the function `ascgen()` (for "*asc generator*"). For example, the chosen cell size of the grid is 500 m (Figure 2.17):

```
> hihi <- ascgen(xy = puechabon$locs[, c("X", "Y")], cellsize = 500)
> image(hihi)
> box()
```

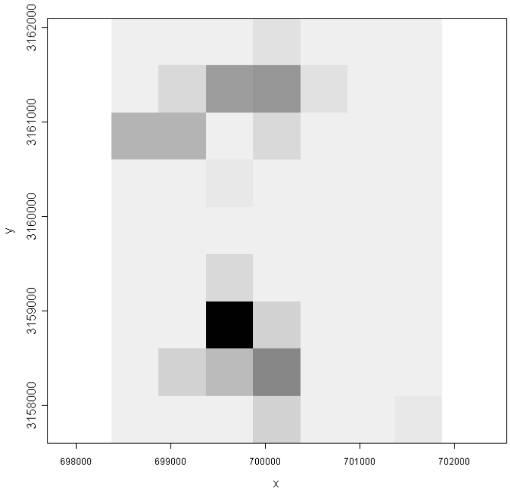


Figure 2.17: The number of wild boars relocations inside each cell of a virtual grid generated with the function `ascgen`.

The function `ascgen()` is useful to define "virtual grids", when no objects of class `"asc"` are available initially. The name of the animals monitored are available in the column `Name` of the object `puechabon$locs`. Once the virtual grid `hihi` has been defined, the number of relocations of each animal falling in each cell of the grid can be drawn using the function `count.points.id()` (Figure 2.18):

```
> tmpbis <- count.points.id(xy = puechabon$locs[, c("X", "Y")],
+     id = puechabon$locs$Name, hihi)
> image(tmpbis)
```

2.3.9 Copying attributes of a map into a matrix with the same dimensions

The function `getascattr()` is extremely useful when one is working with objects of class `"asc"`. For example, we want to divide the elevation of the study area (stored in the object `el` created in section 2.3.2) into two intervals (lower than 200 m a.s.l., and upper than 200 m a.s.l.). This operation can be done by typing the following commands:

```
> el <- getkasc(puechabon$kasc, "Elevation")
> elcat <- el < 200
> class(elcat)
```

```
[1] "matrix"
```

```
> names(attributes(elcat))
```

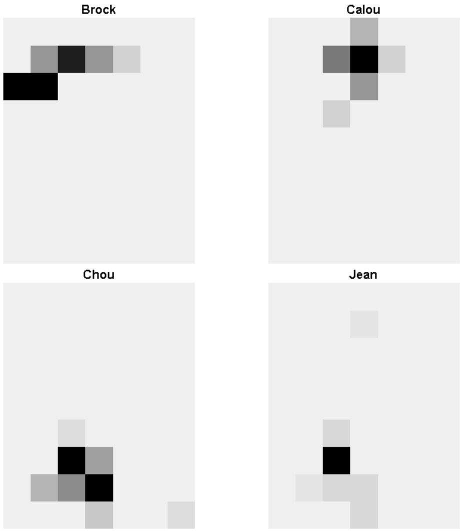


Figure 2.18: The number of wild boars relocations inside each cell of a virtual grid, for each boar.

```
[1] "dim"
```

Note that the result is not an object of class `"asc"`. Further, all the attributes of the map (`cellsize, xll, yll`, see above) have been lost. The function `getascattr()` copies the attributes of an object of class `"asc"` into a matrix of the same dimensions (Figure 2.19):

```
> (elcat <- getascattr(el, elcat, type = "factor", lev = c("> 200 m",
+     "< 200 m")))
```

```
Raster map of class "asc":
Cell size:  100
Number of rows:  121
Number of columns:  111
Type:  factor
```

```
> image(elcat)
> legend(698000, 3165000, levels(elcat), fill = rainbow(2))
```

The function `getkascattr()` is an extension of the function `getascattr()` for objects of class `"kasc"`.

2.3.10 Maps of different areas

It often occurs that the different maps available to an ecologist do not cover exactly the same area (see Figure 2.20). Although the maps have the same dimension, some pixels have values measured for some variables, while other variables are not mapped at this place; the mapped area is not exactly the same for all variables. Most analyses do not deal with these "partially missing pixels". The function `managNAkasc()` sets to `NA` all the pixels that are not mapped for all variables.

The following example renders even clearer this point:

```
> kasc <- puechabon$kasc
> el <- getkasc(kasc, "Elevation")
> sl <- getkasc(kasc, "Slope")
> el[el < 200] <- NA
> tmp <- as.kasc(list(Elevation = el, Slope = sl))
> image(tmp)
```

Note that the two maps in `tmp` do not cover the same area (Figure 2.21);

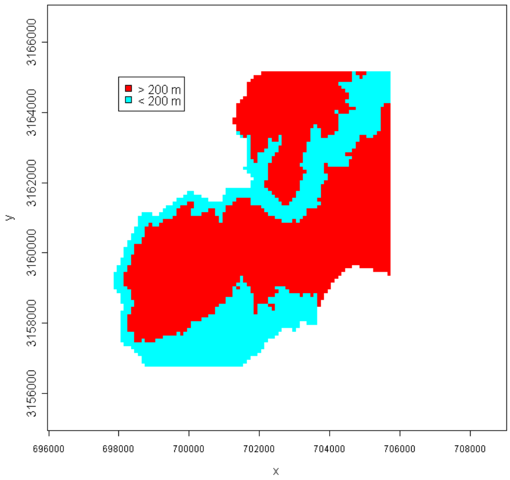


Figure 2.19: The elevation at Puéchabon divided into two classes (< 200 m and > 200 m, see text).

```
> tmp <- managNAkasc(tmp)
> image(tmp)
```

And now, the two maps are covering exactly the same area (Figure 2.22).

2.3.11 The multivariate analysis of objects of class `"kasc"`

As objects of class `"kasc"` store the information about several environmental variables, a multivariate analysis can generate a global view of the area. Actually, the class `"kasc"` is a middle step between the data frames and the maps. However, the object `kasc` contains a lot of missing values (the pixels where the variables are not mapped). The functions `kasc2df()` and `df2kasc()` are intended to convert an object of class `"kasc"` into an object of class `"data.frame"`, and conversely. The following example renders this point clearer:

```
> data(puechabon)
> kasc <- puechabon$kasc
> toto <- kasc[1:10, ]
> class(toto) <- "data.frame"
> toto
```

```
   Elevation Aspect Slope Herbaceous
1         NA   <NA>    NA         NA
2         NA   <NA>    NA         NA
3         NA   <NA>    NA         NA
4         NA   <NA>    NA         NA
5         NA   <NA>    NA         NA
6         NA   <NA>    NA         NA
7         NA   <NA>    NA         NA
8         NA   <NA>    NA         NA
9         NA   <NA>    NA         NA
10        NA   <NA>    NA         NA
```

To have an idea of the structures of the variables on the study area, it is necessary to remove these missing values (cf Figure 2.6 and 2.20), as the functions available in the R package **ade4** do not deal with the missing values:

```
> huhu <- kasc2df(kasc)
> names(huhu)
```

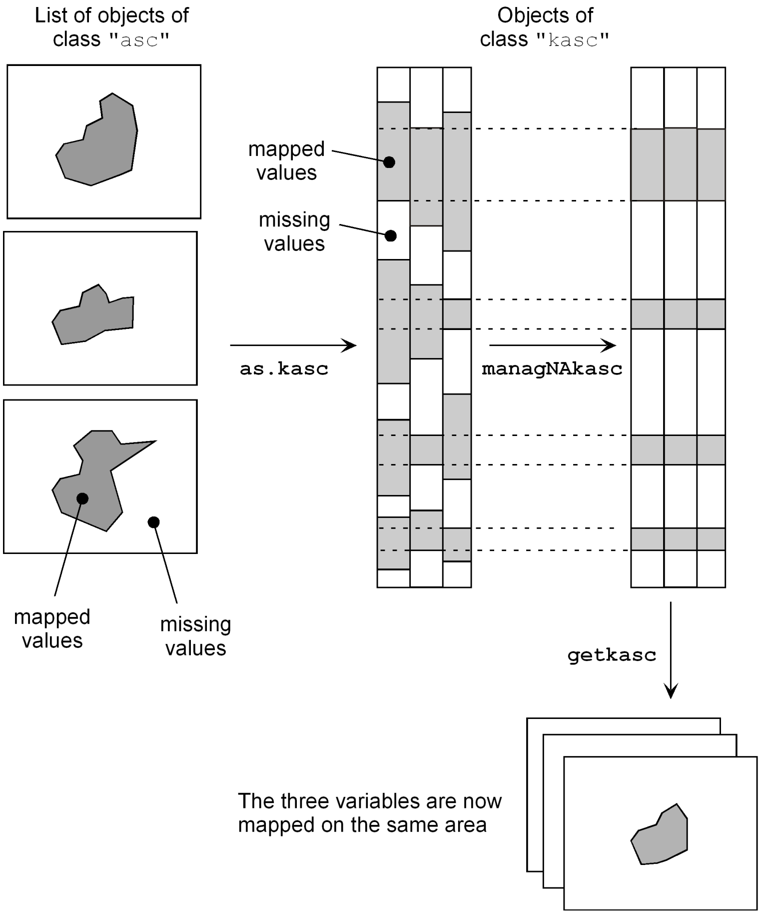


Figure 2.20: The use of `managNAkasc`: A list of objects of class `"asc"` is available, but the areas mapped are not the same for the three variables, even if the maps cover the same area. The function `as.kasc` converts this list into a data frame of class `"kasc"` (see Figure 2.6). Then, the function `managNAkasc` sets to NA all the pixels that are not mapped for all the variables.

Figure 2.21: Two maps (elevation and slope) of Puéchabon stored in the object `tmp`. Note that the two maps do not cover the same area.

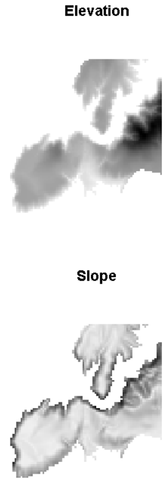


Figure 2.22: The maps of elevation and slope at Puéchabon. Only areas with elevation > 200 m a.s.l. are mapped.

```
[1] "index" "tab"
```

```
> huhu$index[1:4]
```

```
[1] 2017 2018 2019 2020
```

```
> huhu$tab[1:4, ]
```

```
     Elevation    Aspect    Slope Herbaceous
2017        68 SouthWest 1.548994          0
2018        69 SouthWest 1.118594          0
2019        70 SouthWest 1.358634          0
2020        70 SouthWest 1.724311          0
```

The data frame `huhu$tab` contains the pixels for which the mapped variables are not missing, and the vector `huhu$index` contains the indices of the rows of the object `kasc` that are not missing. This vector can be used to back-transform the data frame into an object of class `"kasc"` (see Figure 2.23).

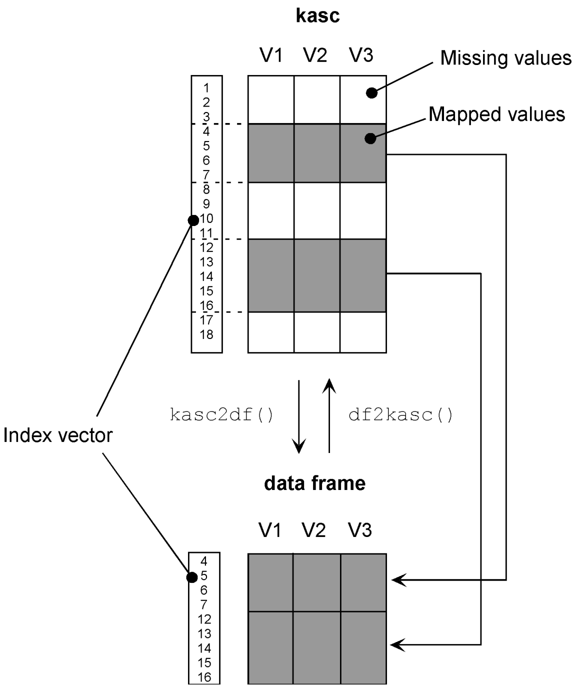


Figure 2.23: The rationale behind the functions `kasc2df` and `df2kasc`.

Once the object of class `"kasc"` has been transformed into a data frame, a multivariate analysis can be done. For example, we keep all the continuous variables (Aspect through out) to compute a Principal Component Analysis:

```
> huhu$tab$Aspect <- NULL
> (pc <- dudi.pca(huhu$tab, scannf = FALSE, nf = 2))
```

```
Duality diagramm
class: pca dudi
$call: dudi.pca(df = huhu$tab, scannf = FALSE, nf = 2)

$nf: 2 axis-components saved
$rank: 3
eigen values: 1.449 0.8475 0.7036
  vector length mode    content
```

```
1 $cw    3      numeric column weights
2 $lw    4379   numeric row weights
3 $eig   3      numeric eigen values

  data.frame nrow ncol content
1 $tab       4379 3    modified array
2 $li        4379 2    row coordinates
3 $l1        4379 2    row normed scores
4 $co        3    2    column coordinates
5 $c1        3    2    column normed scores
other elements: cent norm
```

Then, the first two principal axes can be mapped by the function `df2kasc()`, using as argument the index vector previously returned by the function `kasc2df()` (Figure 2.24):

```
> map <- df2kasc(pc$li, huhu$index, kasc)
> image(map)
```

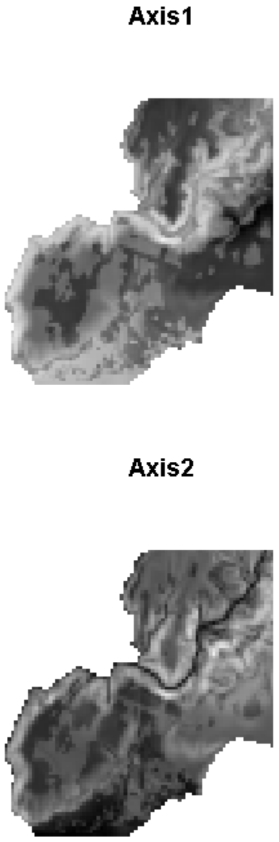


Figure 2.24: Map of the pixels scores in the principal component analysis of the variables elevation, slope and herbaceous cover (Puéchabon, South of France).

2.3.12 Diminishing the resolution of a map

Since R is not a Geographic Information System, it is current that the maps imported into objects of class `asc` or `kasc` are too large, and require too much memory for R . This is the case when high-resolution maps cover a large area. However, a high resolution is not necessarily of major importance in Ecology. Indeed, the ecological studies are often carried out at a given scale, and the precision of the data collected is often linked to this scale. For example, if the distribution of a species is to be studied at the continental scale, it is not essential to have maps with a pixel size of 10 metres. When such cases occur, the function `lowres()` can be used to diminish the resolution of the map. We use here the data set `chamois`, described and loaded in the section 2.3.1:

```
> (kasc <- chamois$map)

Raster map of class "kasc":
Cell size:  50
Number of rows:  353
Number of columns:  353

Variables measured:
1. Vegetation: factor
2. Disteco: numeric
3. Slope: numeric


> (si1 <- object.size(kasc))

[1] 7478324
```

The maps of the area are displayed on Figure 2.25:

```
> image(kasc)
```

Sometimes, we need to get a less precise resolution for a map. For instance, the map `kasc` is rather large and we want to diminish the resolution to 200 m instead of 50 m. So we merge together all pixels contained in a square of 4*4 adjacent pixels of side length 50 m into one pixel of side length 200 m :

```
> (m <- lowres(kasc, np = 4))

Raster map of class "kasc":
Cell size:  200
Number of rows:  88
Number of columns:  88

Variables measured:
1. Vegetation: factor
2. Disteco: numeric
3. Slope: numeric


> image(m)
```

The size of the maps is diminished here:

```
> (si2 <- object.size(m))

[1] 466380

> (si1 - si2)/si1

[1] 0.9376358
```

This results in a large economy of memory (reduction of 93%), without erasing the main features of the area.

2.3.13 Subsetting an area

This operation is somewhat related to the previous one. In some cases, it may be useful to work on a smaller portion of the study area. This can be achieved using the function `subsetmap()`. For example, we subset the map of the Chartreuse mountain, to work on a smaller portion of the map (Figure 2.27):

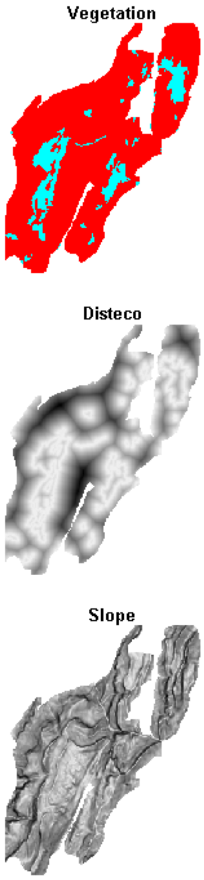


Figure 2.25: Maps of the Chartreuse mountain, in the French Alps (data set `chamois`). The resolution of the map is of 50 m.

Figure 2.26: Same maps as in Figure 2.25, but with a resolution diminished to 200 m.

```
> data(chamois)
> slope <- getkasc(chamois$map, "Slope")
> def.par <- par(no.readonly = TRUE)
> layout(matrix(c(1, 1, 1, 1, 1, 1, 1, 1, 2), ncol = 3, byrow = TRUE))
> par(mar = c(0, 0, 0, 0))
> image(slope, axes = FALSE)
> box()
> x <- c(863603.8, 867286.5)
> y <- c(2042689, 2045797)
> polygon(x = c(x[1], x[2], x[2], x[1]), y = c(y[1], y[1], y[2],
+     y[2]), lwd = 2)
> sl2 <- subsetmap(slope, xlim = x, ylim = y)
> par(mar = c(0, 0, 2, 0))
> image(sl2, axes = FALSE, main = "Reduced map")
> box()
> par(def.par)
```

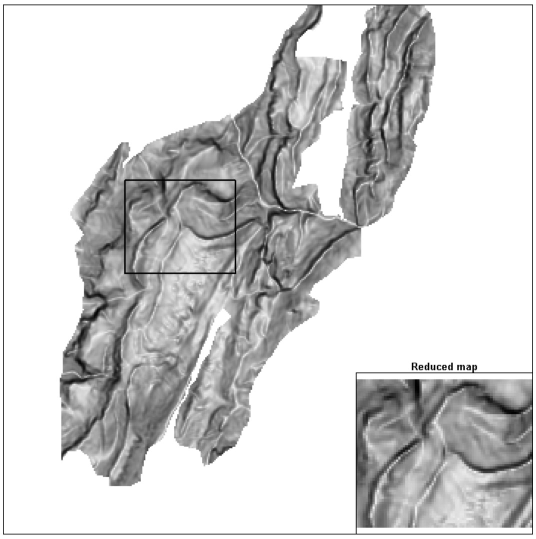


Figure 2.27: The use of the function `subsetmap()` allows to select a part of a given map (delimited by a rectangle), and to store it in another map (labelled "reduced map" in this figure)

In this case, the limits of the new maps were specified in `x` and `y`. Alternatively, when no limits are indicated for the new maps, the user is asked to click on the old map to delimit the upper-left and the bottom-right corners of the new map (try it!). This function can be used on both `"asc"` and `"kasc"` maps.

2.3.14 Other operations

The list presented here does not include all the functions available in **adehabitat** but provides the most frequent basic operations that can be done with the package. Numerous other basic functions are available to the user for programming purposes. For example, the function `labcon()` can be used to label connected features. The function `getXYcoords()` can be used to get the coordinates of the rows and the columns of pixels on the maps. We recommend the user to have a look to the help pages of these functions (type `help.start()`, then choose `packages`, and finally `adehabitat`), and especially to the examples for a better view of what can be done with the package.

Chapter 3

The storage mode of vector maps in adehabitat

3.1 Description of the class `"area"`

The class `"area"` is a class that has been developed in the R package **ade4**. This is the only class of vector maps that can be used with **adehabitat**. The most simple way to create such objects is to use the function `as.area()`. The data set `elec88` of the package **ade4** contains an object that can be converted to this class. The data frame `elec88$area` contains the coordinates of the boundaries of the French departments:

```
> data(elec88)
> ar <- elec88$area
> ar[1:5, ]
```

```
  V1  V2  V3
1 D1 432 213
2 D1 442 199
3 D1 448 204
4 D1 448 219
5 D1 451 227
```

This data frame has three columns. The first variable is a factor defining the polygons. The second and third variables are the x and y coordinates of the polygon vertices in the order where they are found. This format is the standard input to `as.area()`:

```
> ar <- as.area(ar)
> plot(ar)
```

The object `ar` is now a data frame of class `"area"`.

Many functions in **adehabitat** return or require objects of class `"area"`. Thus, the functions `mcp()` (to compute the minimum convex polygon home range), `getverticeshr()` (to compute the kernel home range), or `getcontour()` (to compute the contour polygon of a raster object) return this class of object. This class can also be used in many graphical functions of the R packages **ade4** and **adehabitat**, to overlay vector maps and other types of information (*e.g.* `s.label()`, `plot.traj()`, etc.).

Note: The polygons stored in the object may or may not overlap each other. For example, in the data set `elec88`, displayed previously, the mapped features are the French departments. Since one given location always belongs to only one department, the polygons defined in the object `ar` do not overlap each other. However, in other objects of class `"area"`, the polygons do overlap. For example, let us consider the data set `puechabon`, presented in the section 2.3.1:

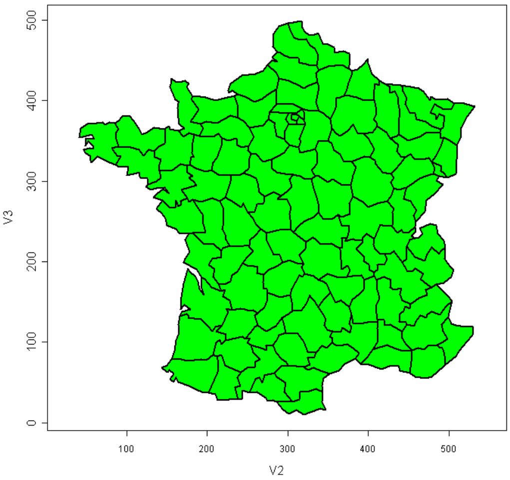


Figure 3.1: Map of France, stored in R as an object of class `"area"`

```
> data(puechabon)
> lo <- puechabon$locs
> cp <- mcp(lo[, c("X", "Y")], lo[, "Name"])
> class(cp)
```

```
[1] "area"       "data.frame"
```

The object `cp` contains the coordinates of the vertices of the home ranges of the four wild boars monitored using radio-tracking, computed as the minimum convex polygon encompassing all relocations. These home ranges may be mapped using the function `plot.area()` (Figure 3.2). Since these animals are not territorial, their home-range overlap each other:

```
> opar <- par(mar = c(0, 0, 0, 0))
> plot(cp, colp = NULL)
> points(puechabon$locs[, c("X", "Y")], pch = 16, col = as.numeric(puechabon$locs$Name))
> box()
> par(opar)
```

It is then possible to export an object of class `"area"` toward a GIS software using the function `area2dxf()`. This function exports an object of class `"area"` to the DXF file format (`http://www.autodesk.com/techpubs/autocad/acad2000/dxf/`). The DXF file format can be read by nearly all GIS softwares. Using the object `cp`, previously defined:

```
> area2dxf(cp, file = "myfile.dxf")
```

This command creates a new file in your working directory, named “myfile.dxf”, that can be read into a GIS. The help page of this function explains how variables measured for each polygon can be exported in the DXF file (*e.g.* exportation of the proportion of tree cover inside each home range).

3.2 Conversion between raster and vector maps

Several functions are available to allow the conversion between raster and vector maps. First, the function `getcontour()` can be used to convert the connected features of an object of class `"asc"` into polygons. Consider the example of the data set `puechabon`:

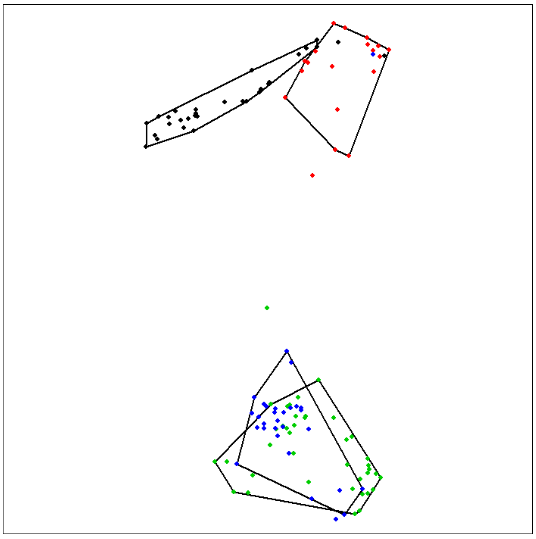


Figure 3.2: Overlapping polygons stored in an object of class area (home ranges of 4 wild boars monitored using radio-tracking at Puéchabon). The relocations of the boars are also displayed.

```
> el <- getkasc(puechabon$kasc, "Elevation")
> cont.el <- getcontour(el)
> class(cont.el)
```

```
[1] "area"       "data.frame"
```

```
> nlevels(cont.el[, 1])
```

```
[1] 1
```

The object `cont.el` is an object of class `"area"`. There is only one connected feature on the map `el`, and this is transformed into one polygon (Figure 3.3):

```
> image(el)
> polygon(cont.el[, 2:3], lwd = 3)
```

Note that more complex maps can also be used. On the other hand, the converse operation (rasterization of vector polygons) is also possible, using the functions `mcp.rast()` and `hr.rast()`. The function `mcp.rast()` converts only one polygon to raster (*i.e.* a data frame with two columns, the x-y coordinates of the vertices of the polygon), whereas the function `hr.rast()` converts an object of class `"area"` to raster. For example, using the data set `puechabon`:

```
> lo <- puechabon$locs
> kasc <- puechabon$kasc
> cp <- mcp(lo[, c("X", "Y")], lo[, "Name"])
```

We define again `cp`, an object of class `area` containing the vertices of the wild boars home ranges (see above). We transform this object into a raster map of class `"kasc"`, using the function `hr.rast()`:

```
> (rast <- hr.rast(cp, kasc))
```

```
Raster map of class "kasc":
Cell size:  100
Number of rows:  121
Number of columns:  111

Variables measured:
1. Brock: numeric
2. Calou: numeric
```

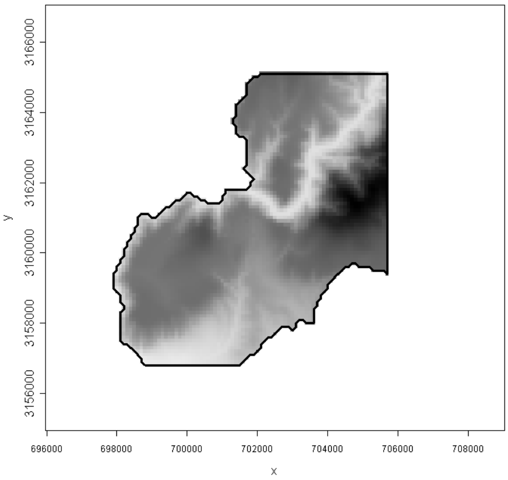


Figure 3.3: The boundaries of the Puéchabon area stored as an object of class `"area"` (computed using the function `getcontour()`). The vertices of the polygon are the centres of the pixels located on the boudary of the mapped area.

```
3. Chou: numeric
4. Jean: numeric
```

```
> def.par <- par(no.readonly = TRUE)
> layout(matrix(c(1, 1, 2, 4, 3, 5), 2, 3))
> par(mar = c(0, 0, 4, 0))
> plot(cp, colp = NULL)
> points(puechabon$locs[, c("X", "Y")], pch = 16, col = as.numeric(puechabon$locs$Name))
> box()
> for (i in names(rast)) {
+     image(getkasc(rast, i), main = paste("Wild boar named", i),
+         axes = FALSE)
+     polygon(cont.el[, 2:3])
+     box()
+ }
> par(def.par)
```

The resulting object is an object of class `"kasc"`, with one column per level of the factor `lo$Name`. Each map in this object corresponds to the (raster) home-range of one animal (Figure 3.4).

3.3 Using masks

The use of masks is a common operation in GIS. A mask allows to specify on a map the limits of a small area, where the subsequent statistical analyses are to be done. The mask is a map of class `"asc"`, that should contain missing values for those areas where processing should not occur and any value for the cells to be processed. We illustrate here an example of use of masks, because it implies knowledge of the functions described previously. A new function, `setmask()`, is introduced here. First, consider the map of elevation at Puechabon and the polygon `pol`, which delimits the plateau (see Figure 3.5):

```
> el <- getkasc(puechabon$kasc, "Elevation")
> pol <- data.frame(x = c(700658, 699222, 698342, 698643, 700427,
+     701029), y = c(3160768, 3160676, 3159402, 3158336, 3158869,
```

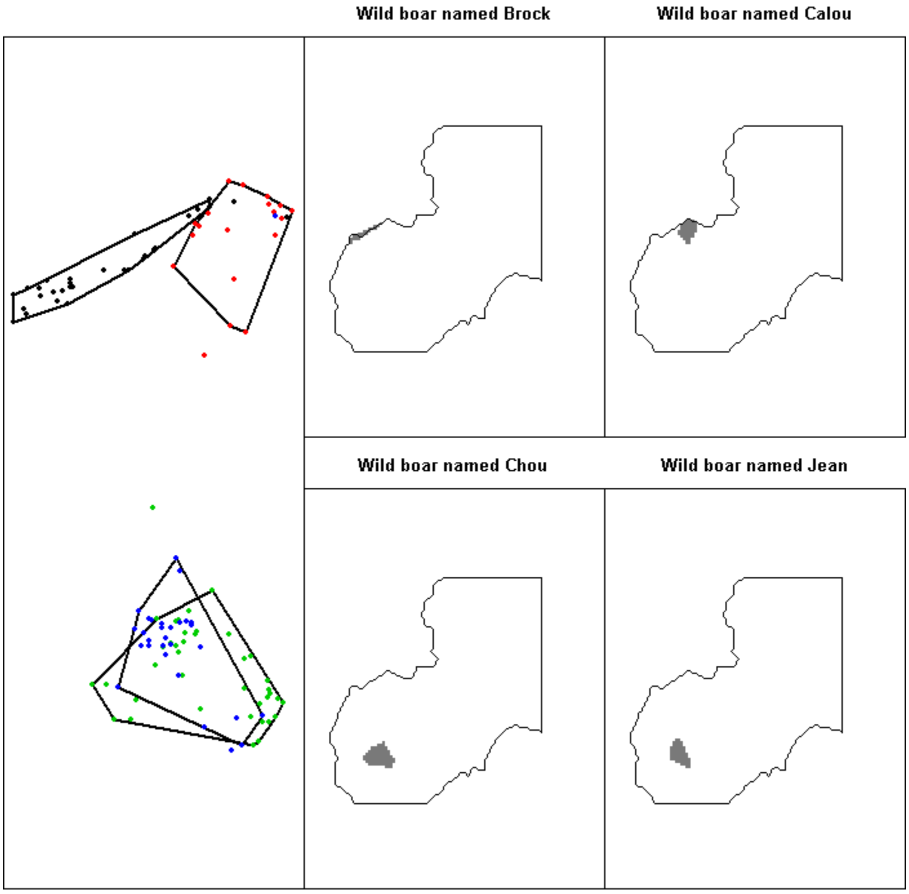


Figure 3.4: left: the home ranges of the four wild boars monitored by radio-tracking, stored as an object of class `"area"`; right: the rasterized home-ranges, stored in an object of class `"kasc"` (computed using the function `hr.rast()`).

```
+      3159657))
> image(el)
> polygon(pol, lwd = 2)
```

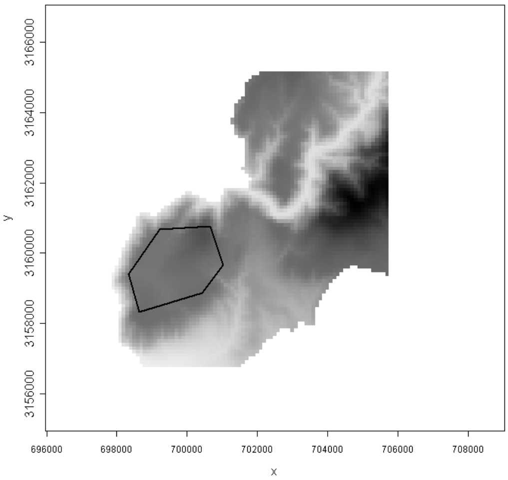


Figure 3.5: The map of elevation at Puéchabon (South of France). The polygon delimits the plateau of Puéchabon.

We rasterize the polygon, to define a mask, and we use the resulting map as argument for the function `setmask()` to keep areas mapped inside the polygon (Figure 3.6):

```
> pr <- mcp.rast(pol, el)
> masked.kasc <- setmask(puechabon$kasc, pr)
> image(masked.kasc, xlim = c(696999, 702373), ylim = c(3156784,
+      3162297))
```

Note that in this case, the coordinates of the masking polygon were available. When the user wants to click on the map to define himself a masking polygon, it is straightforward to write a function using the function `locator()` of the **base** package. For example, the following function is intended to achieve this aim (try it, to defined your own polygon in the example above). The argument `x` is the number of vertices of the polygon:

```
> def.pol <- function(x) {
+      toto <- locator(1)
+      for (i in 2:x) {
+          tutu <- locator(1)
+          toto$x <- c(toto$x, tutu$x)
+          toto$y <- c(toto$y, tutu$y)
+          lines(toto$x, toto$y)
+      }
+      polygon(toto)
+      return(toto)
+ }
```

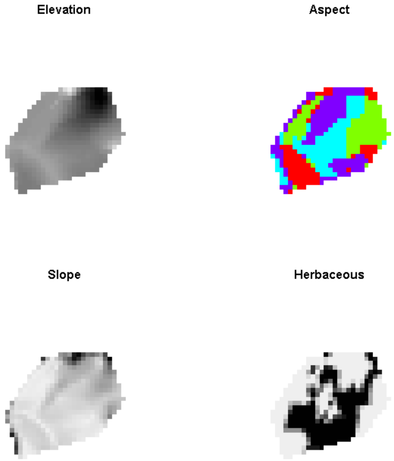


Figure 3.6: The maps of the plateau of Puéchabon (South of France), after the use of the function `setmask`, to mask areas outside the limits of the plateau.

Chapter 4

Conclusion

We have presented here the main classes of map objects available in **adehabitat**. Numerous more specific functions are available to analyse data related to the spatial distribution of animals. These functions concern:

1. the estimation of the home range of animals, using the kernel method (`kernelUD()`), the minimum convex polygon (`mcp()`), the nearest neighbour convex hull (`NNCH()`), or the brownian bridge kernel (`kernelbb()`);
2. the exploration of autocorrelated locations in radio-tracking analysis, using:
   - the computation of turning angles (`angles()`);
   - the computation of travel speeds (`speed()`) ;
   - the computation of the Schoener's ratio (`schoener()`) ;
3. The analysis of habitat selection, using :
   - selection ratios (`wi()`);
   - compositional analysis (`compana()`);
   - habitat suitability maps (algorithm DOMAIN: `domain()`; Mahalanobis distances: `mahasuhab()`; Resource Selection Functions may also be computed with the function `glm()` of the package **base** and the functions presented in this tutorial);
   - multivariate analyses relying on the concept of ecological niche (ENFA: `enfa()`; K-select analysis: `kselect()`; niche analysis, from the package **ade4**: `niche()`).

We encourage the reader look at the help page of these functions.

Note that the package also contains several functions allowing conversion of classes `asc`, `kasc`, `traj` and `area` toward the spatial classes of the package **sp**.

We hope that this package will be useful to biologists concerned by the analysis of the space use by animals.

Chapter 5

Acknowledgements

I warmly thank Anne-Béatrice Dufour and Mathieu Basille (Laboratoire de Biométrie, University of Lyon, France) for their useful comments on earlier drafts of this manuscript. I am also grateful to Mathieu Garel, Gaelle Darmon, Jodie Martin, Bram Van Moorter (Laboratoire de Biométrie, University of Lyon, France), Cyrille Chatelain (Conservatoires et Jardins Botaniques de Genève, Switzerland), Jason Law (University of Michigan, USA) and Sander Oom (University of the Witwatersrand, South Africa) who tested **adehabitat** and advised me about the improvements to bring to the package.

Annexe 14

Didacticiel 2 :

“How to analyze location data”

Clément Calenge, Paolo Cavallini & Carlotta Canova.

Didacticiel de la bibliothèque de fonctions **adehabitat** disponible sur le site du groupe “Animove", URL: http://www.faunalia.com/animov/howto.php

AniMove project
AniMove mailing list
AniMove HowTo
Wiki page
AdeHabitat R Package
GRASS
R
Faunalia

Please help us improving this page by downloading it, modifying it and sending the modified file **here**.

EDITING UNDERWAY!

How to analyze location data

Clément Calenge **calenge@biomserv.univ-lyon1.fr**

Paolo Cavallini **cavallini@faunalia.it**

Carlotta Canova **carlottacanova@tiscali.it**

Step 1: install R

Follow the instructions appropriate for your Operating System; R is packaged for all major GNU/Linux distributions, and for Windows. You can download packages from **R web site**. For Debian you can simply type (as root):

```
apt-get install r-base
```

Step 2: install all necessary libraries

You will need:
ade4
adehabitat
gpclib
Please note: due to frequent updates and improvements, package names change often; in case the links above do not work, please search the packages from **R web site (contributed packages)**.
To install these packages open a shell and write

```
su
```

so you can work as super user writing the root's password. Then you type (for every package)

```
R CMD INSTALL [followed by the path to the package you downloaded,
including the file name]
```

Then exit the root shell.
Another (better? but you need an Internet connection) way of installing R packages is:
start R as superuser, then:

```
install.packages(c("gpclib","ade4","adehabitat"),
dependencies=TRUE, repos="http://cran.cnr.berkeley.edu/")
```

Other mirrors:
http://microarrays.unife.it/CRAN/
http://cran.arsmachinandi.it/
http://microarrays.unife.it/CRAN/
http://rm.mirror.garr.it/mirrors/CRAN/
http://dssm.unipa.it/CRAN/
Please note: some interfaces of R provide easier ways for installing additional packages). See their respective instructions).

Step 3: start R

Start R (specific instruction depend on your Operating System) and load all the relevant libraries:

```
library(ade4)
library(adehabitat)
library(gpclib)
```

Analyses

Now let's play! You can load the sample data from the adehabitat package (dataset *puechabon*), or use your own.

```
data(puechabon)
#loads the data [note: comments to commands start with "#"]
xy<-puechabon$locs[,c("X","Y")]
#takes the columns named X and Y from the section "locs" of the
puechabon dataset, and store them in a matrix named "xy"
id<-puechabon$locs[,c("Name")]
#takes the column named Name from the section "locs" of the
puechabon dataset, and store it in a vector named "id"
#then typing:
> xy
#we can see the value of x and y coordinate
    X     Y
1 699889 3161559
2 700046 3161541
3 698840 3161033
4 699809 3161496
. . . .
#typing
> id
#as you can see, data refer to four animals (Brock Calou Chou Jean)
[1] Brock Brock Brock Brock Brock Brock Brock Brock Brock Brock
Brock Brock
. . .
[37] Calou Calou Calou Calou Calou Calou Calou Calou Calou Calou
Calou Calou
. . .
[61] Chou Chou Chou Chou Chou Chou Chou Chou Chou Chou Chou Chou
. . .
[109] Jean Jean Jean Jean Jean Jean Jean Jean Jean Jean Jean
Levels: Brock Calou Chou Jean
```

To use your data, you can start with a text file (R can also link to a database, use other file formats, etc., in case you need this).
For example to import a file **fagiani** in R you can write:

```
>data<-read.table(file="the path to your file,including the file's
name", header=T, sep=";")
#data is the name that you give to your data
#header=T indicate that in the file's first line there are the
variables' names
#sep=";" means that the columns are separed by ; but if you want,
you can for example, separe the data with the tab writing "\t"
```

You should have a matrix with at least 2 columns, one for x coordinates and one for y; additional columns allow you to run separate analyses (e.g. of different individuals, seasons, activities, etc.).
To remove missing data, just type *(na.omit(yourdata))* when loading data, instead of *(yourdata)*, e.g. xy<-(na.omit(puechabon)). Better do it as a first step, to avoid having misaligned vectors or matrices (e.g. if you have missing xy, but known id).

Summary Statistics

Once you have your data in an R object (we have called it xy), you can start getting some results.

Sample Size:

Total number or records (observations) in the dataset.

```
#Sample Size
nrow(xy)
#per animal
table(id)
```

Minimum X e Y:

Minimum value of x coordinate (northing) and y coordinate (easting).
R commands are:

```
#Minimum X and Y
apply(xy, 2, min)
#Minimum X and Y per animal
apply(xy, 2, function(x) tapply(x, id, min))
```

Mean of X and Y:

Mean value of x coordinate (northing) and y coordinate (easting).

```
#Mean of X and Y
apply(xy, 2, mean)
#Mean of X and Y per animal
apply(xy, 2, function(x) tapply(x, id, mean))
```

XY Variance:

The XY variance (not the covariance).

```
#Variance of X and Y
apply(xy, 2, var)
#Variance of X and Y per animal
apply(xy, 2, function(x) tapply(x, id, var))
```

Distance:

Distance(m) between relocations.

```
di<-as.matrix(dist(xy))
diag(di)<-NA
#Distance between relocations, per animal
xyperan<-split(xy, id)
dibis<-lapply(xyperan, function(x) as.matrix(dist(x)))
dibis<-lapply(dibis, function(x) {diag(x)<-NA;x})
```

Minimum and maximum distance:

Minimum and maximum distance (m) between relocations.

```
min(c(di), na.rm=TRUE)
max(c(di), na.rm=TRUE)
```

Minimum distance:

Minimum distance (m) between relocations.

```
#min distance
lapply(dibis, function(x) min(c(x), na.rm=TRUE))
```

Maximum distance:

Maximum distance (m) between relocations.

```
#max distance
lapply(dibis, function(x) max(c(x), na.rm=TRUE))
```

Distance between successive relocation

To do the following analyses in your data there must be a column date, as in the dataset puechabon.
In a first time you must transform the dataset into an object of class POSIX and then into an object of class traj.
The commands are:

```
da <- as.POSIXct(strptime(as.character(puechabon$locs$Date),
"%y%m%d"))
tr <- as.traj(id = id, xy = xy, date = da)
tr
```

Minimum date

Earliest observation date per dataset.

```
kk<-tapply(tr$date, tr$id, min)
class(kk)<-"POSIXct"
kk
```

Maximum date

Latest observation date per dataset.

```
kk<-tapply(tr$date, tr$id, max)
class(kk)<-"POSIXct"
kk
```

Minimum speed (units/day):

Minimum number of meters traveled per day.

```
#Compute speed
sp<-speed(tr)
#Minimum speed
tapply(sp$speed, sp$id, min)
```

Maximum speed (units/day):

Maximum number of meters traveled per day.

```
#Compute speed
sp<-speed(tr)
#Maximum speed
tapply(sp$speed, sp$id, max)
```

Mean daily speed:

Mean number of meters traveled per day; distance/number of days in dataset.

```
#Compute speed
sp<-speed(tr)
#Mean speed
tapply(sp$speed, sp$id, mean)
```

Distance (m) between successive observation

```
disuc<-sp$speed*sp$dt
```

Total distance (m) per animal

```
ditot<-tapply(disuc, sp$id, sum)
```

Distance (m) between the first and last relocation

```
dd<-function(x) {
xfirst<-x[1,]
```

```
xend<-x[nrow(x),]
di<-sqrt((xfirst[1]-xend[1])^2 + (xfirst[2]-xend[2])^2)
return(di)
}
dilin<-unlist(lapply(xyperan, dd))
```

Mean distance (m) per animal

```
tapply(disuc, sp$id, mean)
```

Linearity

The distance between travel path endpoints and the total distance traveled.

```
dilin/ditot
```

Schoener's ratio

Schoener's ratio for examining autocorrelation.
R2/MSD between successive observations MSD (mean squared distance) is a measure of dispersion of the data.

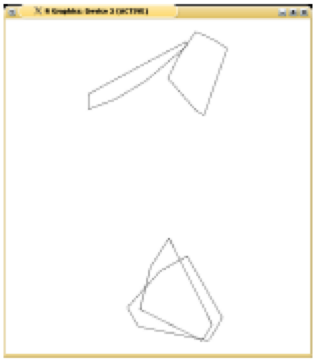
```
t2 <- lapply(split(disuc, sp$id), function(x) sum(x^2)/length(x))
r2 <- lapply(xyperan, function(x) var(x[,1])+var(x[,2]))
ratio <- unlist(t2)/unlist(r2)
```

Home range analyses

Minimum Convex Polygon Home Range

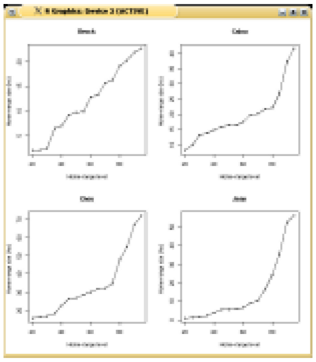
Also called convex hull: calculates a minimum convex polygon home range based on either the selected records or if none are selected the entire data set. A polygon shape file theme is created. The location point statistics menu choice will output a MCP as a graphic object if this is desired rather than a shape file.
If only area is desired then location point statistics with nothing selected will output MCP area.
This calculates area (m) between all points in dataset.

```
hr<-mcp(xy, id)
area.plot(hr)
#The sequent graphic shows the MCP area per animal.
```



```
jj<-mcp.area(xy, id)
jj
   Brock   Calou     Chou     Jean
20 2.01000 8.06000 15.74195 1.13685
25 2.05565 9.96440 16.11330 1.48680
30 2.44080 13.27520 16.67075 1.69070
35 6.46570 13.65330 18.04590 2.33475
40 7.01845 14.88035 22.65340 4.09565
45 9.23435 15.86955 26.36545 5.55715
```

```
50 9.68735 16.31155 26.76340 5.72115
55 9.85540 16.31155 28.51200 5.99155
60 12.84635 17.49680 29.97835 6.64385
65 13.12805 19.74600 31.75860 9.05275
70 15.66545 20.05745 32.25850 10.58390
75 16.14625 21.53025 34.77430 16.40945
80 19.06725 22.15015 47.81925 24.29185
85 20.22535 26.75180 54.76550 35.31225
90 21.82440 37.12975 67.18565 52.59315
95 22.64670 41.68815 71.93205 55.88415
#In this table the first column shows the % of home range and the
others indicate the home range size (ha) for each animal. On the
basis of this one you can obtain the following plot by writing:
plot(jj)
```

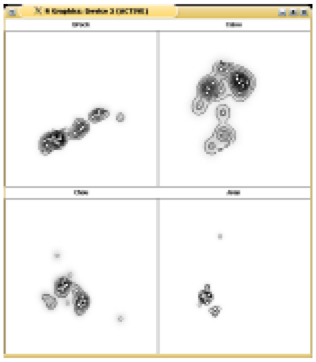


Kernel home range

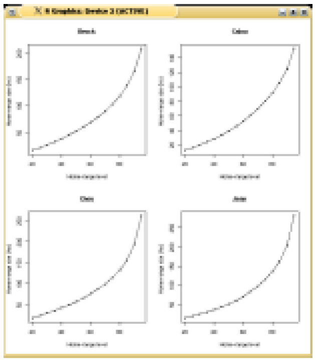
Kernel Home Range calculates a fixed kernel home range utilization distribution (Worton1989) as a grid coverage using either ad hoc calculation of a smoothing parameter, least squares cross validation Silverman (1986), or a user input for the smoothing parameter (H). The bivariate normal density kernel is used as suggested by Worton (1989). The least squares cross validation (LSCV) in this version takes significant processing time despite using the golden minimizing routine. Most user will find that the adhoc calculations are very close to the LSCV for exploratory analysis (for most datasets Hopt is usually between 0 .7 and .9 of Href[Ad hoc] in this implementation). The adhoc calculation is based on Silverman (1986) rather than Worton(1989) or Seaman & Powell (1996). The problem of discretization errors (0 length distances caused from rounding location positions giving a minimized h of zero) are handled slightly different than Tufto (1996). Distance measures are uniform randomly placed between 0 and (href/100) when and only when the distance measurements are 0. This only adjust the locations when necessary and allows for different projection and distance systems. The kernel is based on the non-corrected data. The program queries the user if they would like to adjust the .LSCV or the Adhoc H. Worton (1994) suggests adjusting H by 0.8 for normal distributions and 0.5 for uniform. Work by Seaman & Powel (1996) suggest that this is not necessary with the LSCV. It is our experience that the original Adhoc and LSCV smoothing parameters provide a less biased estimator than a user selected or Worton's corrections.

Three things are output from this routine: A grid coverage with the Utilization Distribution (probabilities), a polygon shapefile containing individual polygons for each selected probability, an associated attribute table containing probability and area fields for each set of probability polygons, and a message box displaying the area calculations of each probability. Note that the default probabilities are 50 and 95, and that the view must be zoomed out sufficiently to encompass the larger probability areas to create the polygon shapefile (95%, etc.).

```
hr<-kernelUD(xy, id, h="LSCV")
hr
image(hr)
#The graphic shows the kernel home range utilization distribution.
```



```
jj<-kernel.area(xy, id)
jj
   Brock    Calou    Chou     Jean
20 15.50560 12.33984 18.82767 14.77481
25 20.35111 15.97852 24.34613 19.20725
30 25.77807 19.85449 30.18920 24.37843
35 31.59267 24.04688 36.35689 30.28835
40 37.79491 28.55566 43.17380 36.19827
45 44.77243 33.53906 49.99072 43.58567
50 52.13760 39.07617 57.78148 50.97308
55 60.27804 45.40430 66.54609 60.57670
60 69.19376 52.68164 76.28454 70.91906
65 78.88477 60.82910 86.99684 83.47765
70 89.93251 69.76758 99.65682 97.51371
75 102.91845 79.73438 114.26450 114.50474
80 118.42406 91.20410 132.11833 135.18947
85 138.19370 105.20508 156.13985 161.78411
90 165.32851 123.79395 194.44442 203.89231
95 209.71331 153.14062 261.63974 281.46004
#In this table the first column shows the % of home range and the
others indicate the home range size (ha) for each animal. On the
basis of this one you can obtain the following plot by writing:
plot(jj)
```



Standard deviation

The standard deviation is computed for each axis (major and minor axis length).

```
## First perform a PCA
pc<-lapply(xyperan, dudi.pca, scale=FALSE, scannf=FALSE)
#Then, the standard deviation is computed for primary and secondary
axes length
sde<-lapply(pc, function(x) apply(x$li, 2, sd))
sde<-do.call("rbind", sde)
```

Eccentricity

The ratio between the minor and the major axis.

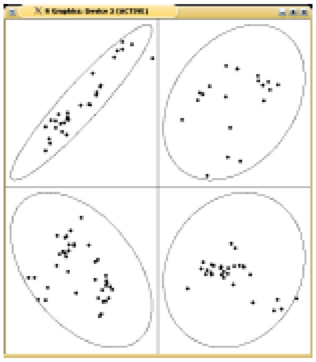
```
apply(sde, 1, function(x) x[2]/x[1])
```

Ellipse

This implements the Jennrich-Turner (1969) Bivariate Normal Home Range. This method of home range calculation is seriously flawed in its dependence on a bivariate normal distribution of locations but is valuable due to its lack of sensitivity with sample size, simplicity of underlying probability theory, ability to get confidence limits on the estimates, and to derive the principle axis of the location coordinates. Other than the rare circumstances where the data is bivariately normal the main utility of this program lies in developing statistically simple home range-habitat models and in comparison with home range estimates using this method in other studies.

The package ellipse contains various routines for drawing ellipses and ellipse-like confidence regions.

```
library(ellipse)
foo<-function(x) {
u<-ellipse(cov(x), centre=apply(x,2,mean))
plot(u, ty="n", axes=FALSE)
points(x, pch=16)
lines(u)
box()
}
par(mfrow=c(2,2), mar=c(0,0,0,0))
lapply(xyperan, foo)
#In this graphic is shown the 95% confidence ellipse:
```



Commands to obtain a file

To convert in a file an object created in R you must write:

```
png("the directory's address in which you want to save the file and
the filename", width = 800, height = 800)
#Obviously you can change width and height as you want.
#Then write the R commands which you have used to create the object
and after
graphics.off()
```

For example, if you want to convert the graphic that shows the MCP area per animal, in a file named MCP, you must write:

```
png("the directory's address in which you want to save the
graphic/MCP", width = 800, height = 800)
area.plot(hr)
graphics.off()
```

GRASS/R Interface

For the integration of R in GRASS, you need to run R from the GRASS shell

enviroment. The interface loads compiled GIS library functions into the R executable enviroment. Then GRASS metadata are transferred to R.
Here we take the result of kernel analysis, we transform it in a shapefile and we import it in GRASS
Installation: connect your computer to the Internet, connect as root, start R and launch the commands:

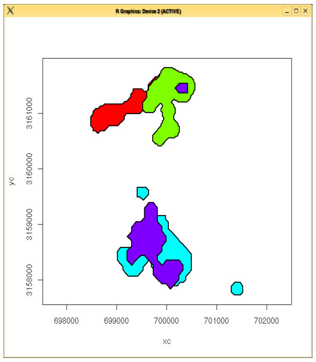
```
install.packages("maptools")
rSpatial <- "http://r-spatial.sourceforge.net/R"
install.packages("spmaptools", repos=rSpatial)
install.packages("spgrass6", repos=rSpatial))
q()
```

Now you exit from root mode. As user you start R and launch the commands:

```
library(spgrass6)
library(spmaptools)
```

We calculate the 95 percent of home range obtained from the command "kernelUD" and plot it

```
ver <- getverticeshr(hr, 95)
plot(ver)
```



"ver" is a list of objects of class "area". There is now a function in adehabitat which allows the conversion to the classes of the package "sp" (see the help page of area2sr). The package "sp" is needed.

```
oo <- lapply(ver, area2sr)
```

"oo" is a list of objects of class "SpatialRings". It is possible to convert each element of this list into a "shapefile" format, but before, we have to convert each object to the class "SpatialRingsDataFrame"

```
kk<-lapply(oo, function(x) SpatialRingsDataFrame(x,
data.frame(name=as.character(1:length(x@polygons)))))
```

Now we create the lists of names of files

```
name <- names(ver)
```

and export to shapefiles

```
lapply(1:length(kk), function(i) writeSRDFShape(kk[[i]], name[i]))
```

The shapefiles are available in your working directory. You can import them in GRASS (see below)

```
q()
```

The function q() finished a R session. When leaving R, you will be asked if you want to save workspace image; If answering Y, the objects are stored within the local directory into the hidden file .RData. When launching R next time in the directory, the objects will be read into the system and you can continue with your work.

Analyses with GRASS

To import a shapefile you can use the line command:

```
v.in.ogr
http://www.grass.itc.it/grass60/manuals/html60_user/v.in.ogr.html
```

To make a minimum convex polygon in GRASS you can use the line command:

```
v.hull
http://www.grass.itc.it/grass60/manuals/html60_user/v.hull.html
```

To make a kernel analyse in GRASS you can use the line command:

```
v.kernel
http://www.grass.itc.it/grass60/manuals/html60_user/v.kernel.html
```

Annexe 15

Manuel d'utilisation :

“Package 'adehabitat' ”

Clément Calenge.

Manuel d'utilisation de la bibliothèque de fonctions **adehabitat**

The adehabitat Package

May 11, 2005

Version 1.3

Date 2005/05/10

Title Analysis of habitat selection by animals

Author Clément Calenge, contributions from Mathieu Basille

Maintainer Clément Calenge <calenge@biomserv.univ-lyon1.fr>

Depends R (>= 1.8.0), ade4

Suggests gpclib

Description A collection of tools for the analysis of habitat selection by animals

License GPL version 2 or newer

R topics documented:

`NNCH` *Nearest Neighbor Convex hull*

Description

`NNCH` computes the home range of several animals using the nearest neighbor convex hull method (Getz and Wilmers, 2004).
`plot.NNCH` gives a graphical display of the estimation.
`NNCH.area` is used for home-range size estimation.
`neighNNCH` is used to determine the optimum number of neighbors (plots the home-range size estimated to a given percentage level according to the number neighbors used for the estimation).
`getverticesNNCH` is used to store the limits of the home ranges.
`plot.NNCHver` displays the limits of the home ranges (but see the section "note").
`NNCH.rast` is used to convert the home range to raster.

Usage

```
NNCH(xy, id = NULL, k = 10, unin = c("m", "km"), unout = c("ha", "km2", "m2"))
plot.NNCH(x, which = names(x), add.points = TRUE, pch = 21,
          bgpts = "white", colpts = "black",
          cex = 0.7, add = FALSE, same4all = TRUE,
          border = NA, ...)
NNCH.area(x, percent = c(95,90,80,70,60,50,40,30,20,10))
neighNNCH(xy, id = NULL, rangek, percent = 95,
          unin = c("m", "km"), unout = c("ha", "km2", "m2"))
plot.ngNNCH(x, ...)
getverticesNNCH(x, percent = 95)
plot.NNCHver<-function(x, which = names(x),
                       colpol = NA,
                       colborder = rep("black", length(x)),
                       lwd = 2, add = FALSE, ...)
NNCH.rast(y, w)
```

Arguments

`xy`	a data frame containing the coordinates of the relocation of the monitored animals
`id`	a factor giving the identity of the animal for each relocation
`k`	the number of neighbors to take into account in the estimation
`unin`	the units of the relocations coordinates. Either `"m"` (default) for meters or `"km"` for kilometers
`unout`	the units of the output areas. Either `"m2"` for square meters, `"km2"` for square kilometers or `"ha"` for hectares (default)
`x`	an object of class "`NNCH`". For `plot.NNCHver`, an object of class `NNCHver` returned by the function `getverticesNNCH`. For `plot.ngNNCH`, an object of class `ngNNCH` returned by the function `neighNNCH`
`which`	a vector of character strings, indicating the names of the animals to be plotted
`add.points`	logical. `TRUE` if the relocations are to be plotted

`pch`	if `add.points` is `TRUE`, the size of the points to be used for the plot of the relocations (see `par`)
`bgpts`	if `add.points` is `TRUE`, the background color to be used for the plot of the relocations (see `par`)
`colpts`	if `add.points` is `TRUE`, the color of the points to be used for the plot of the relocations (see `par`)
`cex`	if `add.points` is `TRUE`, the size of the points to be used for the plot of the relocations (see `par`)
`add`	logical. Whether the home ranges should be displayed on a previous plot
`border`	The color of the border of the polygons. `NA` if a border is not required
`same4all`	logical. If `TRUE`, the plot displaying the home ranges has the same X and Y limits for all animals
`...`	additional arguments to be passed to the generic function `plot`
`percent`	for `NNCH.area`, a vector of percentage levels for home-range size estimation. For `getverticesNNCH`, one value giving the percentage level for the home range estimation.
`colpol`	a vector of the color for filling the polygon. The default, `NA`, is to leave polygons unfilled
`colborder`	a vector of the color to draw the border. The default. Use `border = NA` to omit borders
`lwd`	the border width, a **positive** number
`w`	an object of class `asc` or `kasc`
`rangek`	a vector giving several numbers of neighbors to test

Value

`NNCH` returns a list of class `NNCH`.
`NNCH.area` returns an object of class `hrsize`.
`neighNNCH` returns a matrix of class `ngNNCH`.
`getverticesNNCH` returns an object of class `NNCHver`.
`NNCH.rast` returns a raster map of class `asc` or `kasc`.

Warning

These functions require the package `gpclib`.

Note

The NNCH method can lead to home ranges with holes. In other words, inside the outer limits of the home ranges there are some areas that are not used by the animals. Due to the presence of these holes, it is not possible to stores the home ranges with the class "area", as for the minimum convex polygon (see `?mcp`) or the kernel estimator (see `?kernelUD`). Therefore, the functions of the package gpclib are used here to deal with this kind of home ranges. Note, however, that when the home ranges contain holes, the function `plot.NNCHver` does not displays the holes correctly (filled in with the specified color).

The home range can be estimated at different percentage levels. These levels correspond the percentage of relocations included within the limits of the home range. Note, however, that a given percentage level can correspond to a collection of different home ranges (see example below). Here, we define the NNCH home-range as the *larger* area including the specified percentage.

Finally, note that this function is very long with large datasets.

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

References

Getz, W.M. & Wilmers, C.C. (2004). A local nearest-neighbor convex-hull construction of home ranges and utilization distributions. *Ecography*, **in press**.

See Also

`plot.hrsize` to display the home-range size according to the percentage level of the estimation.

Examples

```
## Not run:

## Can also be used to analyse a point pattern
## For example, locations of chamois in
## the Chartreuse mountains (French Alps)

data(chamois)
xy <- chamois$locs

## choice of the number of neighbors
## Warning: this command can be very long (more than two minutes)
u <- neighNNCH(xy, rangek = 5:30)
plot(u)

## 13 and 20 neighbours seem to be a good choice
## Estimates the home range levels
(aa <- NNCH(xy, k = 13))
plot(aa)

## plot of the home-range size
plot(NNCH.area(aa))

## gets the home range contour
cont <- getverticesNNCH(aa)
plot(cont)

## rasterizes these contours
sl <- getkasc(chamois$map, "Slope")
mm <- NNCH.rast(cont, sl)
image(mm)

## Note that two different sets of polygons can include
## the same percentage of relocation
aa[[1]]$area[97:105,]

## For example, the same percentage of relocations is
## included within the limits of the two home ranges
```

```
## (91.92 percent of the localisations)
opar <- par(mfrow=c(1,2))
plot(aa[[1]]$polygons[[103]],
     poly.args = list(lwd = 2))
points(xy, pch = 16, cex = 0.5)

plot(aa[[1]]$polygons[[98]],
     poly.args = list(lwd = 2))
points(xy, pch=16, cex = 0.5)
par(opar)
## only the left-hand plot is considered as an
## home range estimation (the larger one)

## End(Not run)
```

`angles` *Compute Turning Angles*

Description

`angles` computes the turning angles (in radians) between consecutive moves from an object of class `traj`. See examples for a clearer definition.

Usage

```
angles(x, id = levels(x$id), burst = levels(x$burst),
       date = NULL, slsp =  c("remove", "missing"))
```

Arguments

`x`	an object of class `traj`
`id`	a character vector giving the identity of the animals for which the angles are to be computed
`burst`	a character vector giving the identity of the circuits for which the angles are to be computed (see `as.traj`)
`date`	a vector of class `POSIXct` of length 2 (beginning, end) delimiting the period of interest
`slsp`	a character string. If `"remove"`, successive relocations located at the same place are replaced by a single relocation, allowing the computation of the angles. If `"missing"`, a missing value is returned for the angles when successive relocations located at the same place.

Value

Returns a data frame with the following components:

`id`	the identity of the animal
`x`	the x coordinate of the relocation at which the angle is computed
`y`	the y coordinate of the relocation at which the angle is computed

`date`	a vector of class `POSIXct`, giving the date at which the relocation has been taken
`burst`	the id of the circuit (see `help(as.traj)`)
`angles`	the turning angles between the successive moves.

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

References

Turchin, P. (1998) *Quantitative analysis of movement. Measuring and modeling population redistribution in animals and plants.* Sunderland, Massachusetts: Sinauer Associates.

See Also

`speed` for computation of movement speeds, `as.traj` for additional information about objects of class `traj`

Examples

```
## loads an object of class "traj"
data(puechcirc)
puechcirc

## Gets a part of the trajectory of the wild boar named
## CH93 and draws it
## Also displays the turning angles
toto <- puechcirc[2:5,]
plot(toto$x, toto$y, asp = 1, ylim = c(3158300, 3158550),
     pch = 16,
     main = "Turning angles between\nthree consecutive moves",
     xlab="X", ylab="Y")
lines(toto$x, toto$y)
lines(c(toto$x[2], 700217.6),
      c(toto$y[2], 3158310), lty=2)
lines(c(toto$x[3],700289),
      c(toto$y[3],3158546), lty=2)
ang1x <- c(700234.8, 700231.9, 700231, 700233.7, 700238.8, 700243.2)
ang1y <- c(3158332, 3158336, 3158341, 3158347, 3158350, 3158350)
ang2x <- c(700283.3, 700278.8, 700275.4, 700272.4, 700271.2, 700271.6,
           700274.7)
ang2y <- c(3158522, 3158522, 3158520, 3158517, 3158514, 3158508, 3158504)
lines(ang1x, ang1y)
lines(ang2x, ang2y)
text(700216.1, 3158349, expression(theta[1]), cex=2)
text(700247.7, 3158531, expression(theta[2]), cex=2)
text(c(700301, 700231), c(3158399, 3158487),
     c("Beginning", "End"), pos=4)


## Computation of the turning angles with real data
## on wild boars
plot(puechcirc)
ang <- angles(puechcirc)
```

```
## The angles are in the column Angles:
ang[1:4,]
```

`area2dxf` *Exportation of Areas*

Description

`area2dxf` exports a data frame of class `area` in a DXF file. Such files can be read by virtually all Geographic Information Systems.

Usage

```
area2dxf(x, file, lay = 1:nlevels(factor(x[, 1])))
```

Arguments

`x`	an object of class `area`
`file`	a character string giving the name of the DXF file to be created
`lay`	an optional vector with a length equal to the number of polygons in `x`. This vector is then stored in the field "Layer" of the DXF file (see examples), for GIS mapping

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`as.area`, `area.plot`

Examples

```
## Not run: 
  ## Loads the dataset elec88 from the package ade4
  data(elec88)
  ar <- as.area(elec88$area)
  area.plot(ar, lab = elec88$lab, clab = 0.75)

  ## exportation of the departments toward a dxf file
  area2dxf(ar, file = "Dept", lay = elec88$lab)

  ## Removes the file
  file.remove("Dept.dxf")
## End(Not run)
```

`as.area` *Objects of Class "area"*

Description

Objects of class `area` are used to store the information on the shape of vectorised objects.
An object of class `area` is a data frame with three variables. The first variable is a factor defining the polygons.
The second and third variables are the xy coordinates of the polygon vertices in the order where they are found. This kind of objects are current in the package `ade4`, though this package does not define `area` as a special class.

Usage

```
as.area(x)
```

Arguments

`x`	a data frame with three columns

Value

Returns an object of class `area`

See Also

`area.plot` for other operations on objects of class `area`, `plot.area` to display objects of this class, and `area2dxf` for exportation of the objects `area` toward a GIS.

Examples

```
data(elec88)
ar <- as.area(elec88$area)
plot(ar)
```

`as.kasc` *Working with Several Raster Maps*

Description

`as.kasc` converts a list of matrices of class `asc` into a data frame of class `kasc`.
`getkasc` converts one of the components of a data frame of class `kasc` into a matrix of class `asc`.
`image.kasc` displays a image of maps of class `kasc`.

Usage

```
as.kasc(l)
getkasc(x, var)
image.kasc(x,  var=names(x),
           mar=if (length(var)>1) c(0,0,2,0) else c(5.1,4.1,4.1,2.1),
           axes=(length(var) == 1),
           clfac=NULL, col=gray((240:1)/256), mfrow=NULL, ...)
print.kasc(x, ...)
```

Arguments

`l`	a list of objects of class `asc`
`x`	an object of class `kasc`
`var`	for `getkasc`, a character string or a column number. For `image.kasc`, either a vector of character strings, giving the name of one (or several) variable(s) in x, or a vector of integers, giving the number of the columns to be displayed.
`mar`	this parameter is passed as the parameter `mar` of the function `par` (see `help(par)`). By default, this parameter depends of the number of variables in `x`
`axes`	logical. If `TRUE`, axes are drawn on the maps.
`clfac`	an optional list of vectors of character strings. Each element of the list is a vector corresponding to one factor in `x`. Each element of the vector gives the color name associated with each level of the corresponding factor (see Examples).
`col`	a character vector. For numeric maps, the colors to be used
`mfrow`	A vector of the form `c(nr, nc)`. Subsequent figures will be drawn in an `nr`-by-`nc` array on the device by rows
`...`	additional parameters to be passed to the generic functions `image` and `print`

Value

`as.kasc` returns an object of class `kasc`. The matrices of class `asc` are converted into vectors and coerced into a data frame of class `kasc`. This data frame has the following attributes:

`xll`	the X-coordinate of the center of the lower left pixel of the map
`yll`	the Y-coordinate of the center of the lower left pixel of the map
`cellsize`	the size of a pixel
`nrow`	the number of rows of the map. Caution ! the number of rows of the raster map is the number of columns of the matrix of class `asc`
`ncol`	the number of columns of the map. Caution ! the number of columns of the raster map is the number of rows of the matrix of class `asc`

The different maps in the input list of objects of class `asc` should describe the same area (same attributes: same `cellsize`, same `xll`, same `yll`, and same dimensions).

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`import.asc` for additionnal information on objects of class `asc`

Examples

```
data(puechabon)

## kasc is a list of class kasc
(kasc <- puechabon$kasc)

## Stores two elements of the kasc in objects
## of class asc
(asp <- getkasc(kasc, "Aspect"))
(sl <- getkasc(kasc, "Slope"))

## Displays the slopes on the area (numeric)
image(sl, main = "Aspect", xlab = "Lambert X", ylab = "Lambert Y")

## Aspect is a factor:
## cl is the list of color associated with the levels
## of asp
cl <- colasc(asp, NorthEast = "blue", SouthEast = "red",
             SouthWest = "orange", NorthWest = "green")
image(asp, clfac = cl, main = "Aspect", xlab = "Lambert X",
      ylab = "Lambert Y")
legend(706500, 3162000, legend=levels(asp), fill=cl, cex=0.7)


## Creation of a new kasc with elevation
## and slopes
cuicui <- as.kasc(list(Slope = sl, Aspect = asp))

## Displays the kasc object
## with random colors for aspect and grey levels for slopes
image(cuicui)

## with cm.colors for slopes (numeric)
## and cl for aspect (factor)
image(cuicui, col = cm.colors(256), clfac = list(Aspect = cl))

## plots only slope
image(cuicui, "Slope", main="Slope")
## similar to
image(cuicui, 1, main = "Slope")
```

`as.sahrlocs` *Exploratory Analysis of Habitat Selection*

Description

`as.sahrlocs` creates objects of class `sahrlocs`. This class has a central place in habitat selection studies relying on radio-tracking data. Niche analysis and K-select analysis can be performed using this class of objects. This class may also be used for exploratory purposes. This class of object has three main components: an object of class `kasc` describing the study area, an object of class `kasc` describing what is available to the animals, and an object of class `kasc` describing the relocations of the animals (`sahrlocs` = Study Area - Home Range - reLOCationS). `getsahrlocs` converts one of the components of an object of class `sahrlocs` into an object of class `kasc`.

Usage

```
as.sahrlocs(mlocs, mhr, msa, descan = NULL)
getsahrlocs(x, what = c("sa", "hr", "locs"))
print.sahrlocs(x, ...)
```

Arguments

`mlocs`	an object of class `kasc` returned by the function `count.points.id`
`mhr`	an object of class `kasc` returned by the function `hr.rast`, by the function `buffer.ani`, or by any other user-defined function
`msa`	an object of class `kasc` describing the study area
`descan`	an optionnal data frame with the number of rows equal to the number of monitored animals. Each column of this data frame gives any type of information on the monitored animals (e.g. sex, age, and so on)
`what`	a character string giving the component of the `sahrlocs` object to be converted. Should be either `"sa"` (study area), `"hr"` (home ranges) or `"locs"` (relocations)
`x`	an object of class `sahrlocs`
`...`	additionnal parameters to be passed to the generic function `print`

Details

The different maps in the input list of objects `kasc` should describe the same area (same attributes: same `cellsize`, same `xll`, same `yll` and same dimensions).

Value

Returns one object of class `sahrlocs`, which is a list containing the input arguments for the function `as.sahrlocs`. Objects of class `sahrlocs` have the same attributes as objects of class `kasc` (`xll`, `yll`, `cellsize`, `nrow`, and `ncol`).

`getsahrlocs` returns an object of class `kasc` (see `as.kasc`).

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`as.kasc` for additionnal information on objects of class `kasc`, `hr.rast` and `buffer.ani` for the creation of the `"hr"` component of this object, `count.points.id` for the creation of the `"locs"` component of this object, `plot.sahrlocs` and `image.sahrlocs` for a graphical display of such objects.

Examples

```
data(puechabon)
kasc <- puechabon$kasc
locs <-  puechabon$locs

## Computes the home ranges of the animals...
cp <- mcp(locs[,4:5], locs[,1])
```

```
## ... and converts it to raster
cprast <- hr.rast(cp, kasc)
locrast <- count.points.id(locs[,4:5], locs[,1], kasc)

## Creation of the sahrlocs object
(sahr <- as.sahrlocs(locrast, cprast, kasc))

## adds information on the monitored animals
age <- factor(tapply(locs[,2], locs[,1], mean))
sex <- factor(tapply(locs[,3], locs[,1], mean))
info <- as.data.frame(cbind(sex, age))
(sahr <- as.sahrlocs(locrast, cprast, kasc, info))

## Gets the "study area" component of the object
toto <- getsahrlocs(sahr)
image(toto)
```

`as.traj` *Working with Trajectories in 2D Space*

Description

The class `traj` is intended to explore trajectories of animals monitored using radio-tracking.
`as.traj` creates an object of this class.
`summary.traj` returns the number of relocations for each "burst" of relocations and each animal.
`plot.traj` allows various graphical displays of the trajectories.
`getburst` returns an object of class `traj` satisfying the specified criteria (selection of one focus animal, of a period of interest, of special "bursts" (see details)).
`traj2df`, and the reciprocal function `df2traj` respectively converts an object of class `traj` to an object of class `data.frame`, and conversely.

Usage

```
as.traj(id, xy, date, burst = id, ...)
print.traj(x, ...)
summary.traj(object, id = levels(object$id), date = NULL, ...)
plot.traj(x, id = levels(x$id), burst = levels(x$burst), date = NULL,
          asc = NULL, area = NULL,
          xlim = range(x$x), ylim = range(x$y),
          colasc = gray((256:1)/256), colpol = "green",
          addpoints = TRUE, addlines = TRUE,
          perani = TRUE, final = TRUE, ...)
getburst(x, burst = levels(x$burst),
         id = levels(x$id), date = NULL)
traj2df(x)
df2traj(df)
```

Arguments

`id`	a factor giving for each relocation the identity of the individual monitored in `as.traj`.

	a character vector containing the identity of the individuals of interest in other functions
`xy`	a data frame containing the coordinates of the relocations
`date`	a vector of class `POSIXct` giving the date for each relocation in `as.traj`. a vector of class `POSIXct` with length 2, indicating the beginning and the end of the period of interest in other functions
`burst`	a factor giving the identity of each "burst" of relocations in `as.traj` (e.g. the circuit id, see details). The burst level needs to be unique (two animals cannot have the same burst levels). a character vector containing the burst levels of interest in `plot.traj` and `getburst`
`x`	an object of class `traj`
`object`	an object of class `traj`
`asc`	an object of class `asc`
`area`	an object of class `area` (see `help(area)`)
`xlim`	the ranges to be encompassed by the x axis
`ylim`	the ranges to be encompassed by the y axis
`colasc`	a character vector giving the colors of the map of class `asc`
`colpol`	a character vector giving the colors of the polygon contour map, when `area` is not `NULL`
`addlines`	logical. If `TRUE`, lines joining consecutive relocations are drawn
`addpoints`	logical. If `TRUE`, points corresponding to each relocation are drawn
`perani`	logical. If `TRUE`, one plot is drawn for each level of the factor `id`, and for a given animal, the several bursts are superposed on the same plot. If `FALSE`, one plot is drawn for each level of the factor `burst`
`final`	logical. If `TRUE`, the initial and final relocations of each burst are indicated in blue and red, respectively
`df`	a data frame to be converted to the class `traj`
`...`	other optional vectors containing some variables measured at each relocation (e.g. temperature, wind, elevation, etc.) in `as.traj`. For other functions, arguments to be passed to the generic functions `plot`, `summary` and `print`

Details

For a given individual, trajectories are often sampled as "bursts" of relocations (Dunn and Gipson, 1977). For example, when an animal is monitored using radio-tracking, the data may consist of several circuits of activity (two successive relocations on one circuit are often highly autocorrelated, but the data from two circuits may be sampled at long intervals in time). These bursts are indicated by the factor `burst`.

Value

An object of class `traj` is a data frame with one column named `id`, one column named `x`, one column named `y`, one column named `date` and one column named `burst`. This class therefore inherits from the class `data.frame`.

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

References

Dunn, J.E. and Gipson, P.S. (1977) Analysis of radio telemetry data in studies of home range. *Biometrics*. **59**, 794–800.

See Also

`as.POSIXct` and `strptime` for additional information of the class `POSIX`.

Examples

```
data(puechabon)
locs <- puechabon$locs
locs[1:4,]

### Conversion of the date to the format POSIX
da <- as.character(locs$Date)
da <- as.POSIXct(strptime(as.character(locs$Date),
               "%y%m%d"))

### Creation of the object of class "traj"
(tr <- as.traj(id = locs$Name, xy = locs[,c("X", "Y")],
               date = da))
summary(tr)
plot(tr)

### Displays on maps of the study area
k <- puechabon$kasc
ele <- getkasc(k, "Elevation")
plot(tr, asc = ele)

### Case with several burst per animal
data(puechcirc)
plot(puechcirc, asc = ele)
plot(puechcirc, asc = ele, perani = FALSE)
```

`asc2im` *Conversion of Maps of Class 'asc' and 'im' (Package spatstat)*

Description

These functions convert maps of class `asc` to objects of class `im` (package spatstats) and conversely.

Usage

```
asc2im(x)
im2asc(x)
```

Arguments

`x`	an object of class `asc` or `im`

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`import.asc` for additionnal information on objects of class `asc`, and `im` for additionnal information on objects of class `im`

Examples

```
if (require(spatstat)) {

#######################
### Conversion asc -> im

  data(puechabon)
  el <- getkasc(puechabon$kasc, "Elevation")
  image(el, main = "An object of class \"asc\"")
  elim <- asc2im(el)
  image(elim, main = "An object of class \"im\"")

#######################
### Conversion im -> asc
  u <- matrix(rnorm(10000), 100, 100)
  haha <- im(u)
  image(haha, main = "class im")
  hihi <- im2asc(haha)
  image(hihi, main = "class asc")

}
```

`ascgen` *Creation of Raster Maps*

Description

`ascgen` creates an object of class `asc` using a set of points.

Usage

```
ascgen(xy = NULL, cellsize = NULL, nrcol = 10, count = TRUE)
```

Arguments

`xy`	a data frame with two columns: the x and y coordinates of the points
`cellsize`	the `cellsize` attribute of the object of class `asc` to be created
`nrcol`	the size of the square raster map to be created (number of rows and columns)
`count`	logical. If `TRUE`, the object of class `asc` contains the number of points in each cell. If `FALSE`, all the cells are set to zero

Value

Returns an object of class `asc`.

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`import.asc` for additional information on objects of class `asc`.

Examples

```
data(puechabon)
lo <- puechabon$locs[,c("X","Y")]
plot(lo, asp = 1, pch = 16)

## lo contains the relocations of wild boars
rast <- ascgen(lo, cellsize = 100)
image(rast)

## Alternatively, one can specify the size of the square raster map
rast <- ascgen(lo, nrcol = 10)
rast
image(rast)

## can be used for further analyses
## (e.g. correspondence analyses)
locs <- puechabon$locs[, c("Name", "X", "Y")]
o <- count.points.id(locs[,2:3], locs[,1], rast)
image(o)
```

`bighorn` *Radio-Tracking of Bighorn Sheeps*

Description

This data set describes the habitat use and availability for 6 bighorn sheeps monitored by radio-tracking (Arnett et al. 1989, in Manly et al., 2003, p. 67-74). 10 habitat types are considered.

Usage

```
data(bighorn)
```

Details

The object `bighorn` is a list, with the following components:

- **`used`** the number of resource units used by each animal (in rows) in each habitat category (in columns).
- **`availTrue`** the availability of each habitat category.
- **`availEstimated`** a sample of available resource units in each habitat category.

References

Manly, B.F.J., McDonald, L.L., Thomas, D.L., McDonald, T.L. & Erickson, W.P. (2003) *Resource selection by animals - Statistical design and Analysis for field studies. Second edition.* London: Kluwer academic publishers.

`biv.test` *Bivariate Test*

Description

`biv.plot` displays a bivariate plot. `biv.test` displays the results of a bivariate randomisation test.

Usage

```
biv.plot(dfxy, br = 10, points = TRUE, density = TRUE,
         kernel = TRUE, o.include = FALSE, pch, cex, col, h, sub,
         side = c("top", "bottom", "none"), ...)
biv.test(dfxy, point, br = 10, points = TRUE, density = TRUE,
         kernel = TRUE, o.include = FALSE, pch, cex, col, Pcol, h, sub,
         side = c("top", "bottom", "none"), ...)
```

Arguments

`dfxy`	a data frame with N lines (couples of values) and two columns
`br`	a parameter used to define the numbers of breaks of the histograms. A larger value leads to a larger number of breaks
`points`	logical. Whether the points should be displayed
`density`	logical. Whether the kernel density estimation should be displayed for the marginal histograms
`kernel`	logical. Whether the kernel density estimation should be displayed for the bivariate plot
`o.include`	logical. If `TRUE`, the origin is included in the plot
`pch`	plotting "character", i.e., symbol to use for the points. (see `?points`)
`cex`	character expansion for the points
`col`	color code or name for the points, see `?par`
`h`	vector of bandwidths for x and y directions, used in the function `kde2d` of the package `MASS`. Defaults to normal reference bandwidth (see `?kde2d`)
`sub`	a character string to be inserted in the plot as a title
`side`	if `"top"`, the x and y scales of the grid are upside, if `"bottom"` they are downside, if `"none"` no legend
`point`	a vector of length 2, representing the observation to be compared with the simulated values of the randomisation test
`Pcol`	color code or name for the observed point
`...`	further arguments passed to or from other methods

Details

`biv.test` is used to display the results of a bivariate randomisation test. An example of use of the function is provided in the function `niche.test`.

The x-axis of the main window corresponds to the first column of `dfxy`; the y-axis corresponds to the second column. Kernel density is estimated to indicate the contours of the distribution of randomised values. The two marginal histograms correspond to the univariate tests on each axis, for which the p-values are computed with `as.randtest` (one-sided tests).

Warning

`biv.plot` and `biv.test` uses the function `kde2d` of the package `MASS`.

Author(s)

Mathieu Basille ⟨basille@biomserv.univ-lyon1.fr⟩

See Also

`as.randtest`, `niche.test`

Examples

```
## Not run:
x = rnorm(1000,2)
y = 2*x+rnorm(1000,2)
dfxy = data.frame(x, y)

biv.plot(dfxy)
biv.plot(dfxy, points=F, col="lightblue", br=20)

p = c(3, 4)
biv.test(dfxy, p)
biv.test(dfxy, p, points=F, Pcol="darkred", col="lightblue", br=20)
## End(Not run)
```

`buffer` *Compute Buffers*

Description

`buffer` compute buffers around a set of locations.
`buffer.ani` is to be used when the points can be grouped into several categories (e.g. the relocations of several animals monitored using radio-tracking; the function `buffer` is then applied to each animal).
`buffer.line` compute buffers around a line.

Usage

```
buffer(pts, x, dist)
buffer.ani(pts, fac, x, dist)
buffer.line(xy, x, dist)
```

Arguments

`pts`	a data frame with two columns (x and y coordinates of the points)
`x`	either an object of class `asc` or `kasc` with the same attributes as those desired for the output, or an object of class `mapattr` (see `storemapattr`)
`dist`	a value of distance
`fac`	a factor defining the categories of the points
`xy`	a data frame containing the coordinates of the vertices of the lines

Value

`buffer` and `buffer.line` return an object of class `asc`, with 1 for pixels located within a specified distance of given points, and `NA` otherwise.
`buffer.ani` returns a data frame of class `kasc`, with each column corresponding to one level of the factor `fac`.

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`as.kasc` for additionnal information on objects of class `kasc`, `import.asc` for further information on objects of class `asc`, `storemapattr` for further information on objects of class `mapattr`.

Examples

```
data(puechabon)

# locs is the data frame containing the
# relocations of wild boars monitored by radio-tracking
locs <- puechabon$locs

# sa is the "kasc" object of maps of the study area
sa <- puechabon$kasc

# Buffer of 500 m around all relocations
bu <- buffer(locs[,4:5], sa, 500)
image(bu)

# displays all the pixels of the study area within 500 m
# of a relocation of each monitored wild boar
buani <- buffer.ani(locs[,4:5], locs[,1], sa, 500)
image(buani)

## buffer around a trajectory
data(puechcirc)
uu <- getburst(puechcirc, burst = "CH930824")
w <- ascgen(uu[,c("x","y")], nrcol = 100)
out <- buffer.line(uu[,c("x","y")], w, 100)
image(out)
plot(uu, asc = out)
```

`chamois` *Location of Chamois Groups in the Chartreuse Mountains*

Description

This data set describes the habitat use and availability by the chamois of the Chartreuse mountains (Isère, France), in 1992 and 1997. These data have been gathered during the hunting season (Fall).

Usage

```
data(chamois)
```

Details

The object `chamois` is a list containing the following components:

`locs` A data frame containing the x and y coordinates of 198 chamois groups.

`map` a map of class `kasc` describing the vegetation (Forest or Open areas), the distance from the ecotone Open/Forest and the slopes on the area.

References

Fédération Départementale des Chasseurs de l'Isère, 65 av Jean Jaurès, 38320 Eybens. France.

`colasc` *Creates a Vector of Colors for a Raster Map of Type 'factor'*

Description

`colasc` creates a vector of colors for a raster map of class `asc` and of type `"factor"`.

Usage

```
colasc(x, ...)
```

Arguments

`x`	an object of class `asc`.
`...`	arguments named as the levels of the factor, with character values equal to the colors for these levels (see examples)

Value

Returns a character vector.

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`import.asc`

Examples

```
data(puechabon)

## gets the aspect in an asc object
asp <- getkasc(puechabon$kasc, "Aspect")

## creates the vector of colors
cl <- colasc(asp, NorthEast = "blue", SouthEast = "red",
             SouthWest = "orange", NorthWest = "green")

## graphical display
image(asp, clfac = cl, main = "Aspect", xlab = "Lambert X",
      ylab = "Lambert Y")
legend(706500, 3162000, legend=levels(asp), fill = cl, cex = 0.7)
```

`compana` *Compositional Analysis of Habitat Use*

Description

`compana` performs a classical compositional analysis of habitat use (Aebischer et al., 1993).

Usage

```
compana(used, avail, test = c("randomisation", "parametric"),
        rnv = 0.01, nrep = 500, alpha = 0.1)
```

Arguments

`used`	a matrix or a data frame giving the percentage of use of habitats (in columns) by animals (in rows).
`avail`	a matrix or a data frame describing the percentage of availability of habitats (in columns) by animals (in rows).
`test`	a character string. If `"randomisation"`, randomisation tests are performed for both the habitat ranking and the test of habitat selection. If `"parametric"`, usual parametric tests are performed (chi-square for the test of habitat selection and t-tests for habitat ranking).
`rnv`	the number replacing the 0 values occurring in the matrix `used`.
`nrep`	the number of repetitions in the randomisation tests.
`alpha`	the alpha level for the tests.

Details

The compositional analysis of habitat use has been recommended by Aebischer et al. (1993) for the analysis of habitat selection by several animals, when the resources are defined by several categories (e.g. vegetation types).
This analysis is carried out in two steps: first the significance of habitat selection is tested (using a Wilks lambda). Then, a ranking matrix is built, indicating whether the habitat type in row is significantly used more or less than the habitat type in column. When this analysis is performed on radio-tracking data, Aebischer et al. recommend to study habitat selection at two scales: (i) selection of the home range within the study area, and (ii) selection of the relocations within the home range. The first scale is termed second-order habitat selection by Johnson (1980), and the second, third-order habitat selection.

When zero values are found in the matrix of used habitats, they are replaced by a small value (by default, 0.01), according to the recommendations of Aebischer et al. (1993).

When zero values are found in the matrix of available habitats, the function `compana` uses the procedure termed "weighted mean lambda" described in Aebischer et al. (1993: Appendix 2), instead of the usual lambda (see examples). Zero values can be found in the matrix of available habitats when the third-order habitat selection is under focus. In this case, it may occur that some habitat types are available to some animals and not to the others.

Value

Returns a list of the class `compana`:

`used`	the matrix of used habitats
`avail`	the matrix of available habitats
`type.test`	a character string. Either `"randomisation"` or `"parametric"`
`test`	the results of the test of habitat selection
`rm`	the ranking matrix: a square matrix with `nh` rows and `nh` columns, where `nh` is the number of habitat types under study. At the intersection of the row `i` and of the column `j`, there is a `"+"` when the habitat `i` is more used than the habitat in column, and `"-"` otherwise. When the difference is significant, the sign is tripled.
`rmnb`	the matrix containing the number of animals used to perform the tests in `rm`.
`rank`	the rank of the habitat types. It is equal to the number of `"+"` for each habitat type in row of `rm`.
`rmv`	the matrix of statistics used to build `rm`. if `(test = "parametric")`, the matrix contains the values of `t`, in the t-test comparing the row and the column habitat. if `(test = "randomisation")`, the matrix contains the mean difference between the used and available log-ratios (see Aebischer et al., 1993).
`profile`	The profile of preferences: resource types are sorted so that the left type is the most preferred and the right type is the most avoided. Habitats for which the intensity of habitat selection is similar (no significant difference) are connected by a line.

Note

In the examples below, the results differ from those published in Aebischer et al. (squirrel example, selection of the relocations within the home range). In fact, there has been a confusion in the column names in the paper. Actually, Aebischer (pers. com.) indicated that the ranking matrix given in this example is correct.

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

References

Aebischer, N. J. and Robertson, P. A. (1992) Practical aspects of compositional analysis as applied to pheasant habitat utilisation. pp. 285–293 In: Priede, G. and Swift, S. M. *Wildlife telemetry, remote monitoring and tracking of animals*.

Aebischer, N. J., Robertson, P. A. and Kenward, R. E. (1993) Compositional analysis of habitat use from animal radiotracking data. *Ecology*, **74**, 1313–1325.

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

See Also

`wiII` and `wiIII` for other analyses of habitat selection with this kind of data.

Examples

```
## The examples presented here
## are the same as those presented in
## the paper of Aebischer et al. (1993)

#############################
## Pheasant dataset: first
## example in Aebischer et al.

data(pheasant)

## Second order habitat selection
## Selection of home range within the
## study area (example of parametric test)
pheana2 <- compana(pheasant$mcp, pheasant$studyarea,
                   test = "parametric")
pheana2

## The ranking matrix:
print(pheana2$rm, quote = FALSE)

## Third order habitat selection
## (relocation within home range)
## We remove the first pheasant of the analysis
## (as in the paper of Aebischer et al.)
## before the analysis
pheana3 <- compana(pheasant$locs[-1,], pheasant$mcp[-1,c(1,2,4)])
pheana3

## The ranking matrix:
print(pheana3$rm, quote = FALSE)
```

```
#############################
## Squirrel data set: second
## example in Aebischer et al.

data(squirrel)

## Second order habitat selection
## Selection of home range within the
## study area
squiana2 <- compana(squirrel$mcp, squirrel$studyarea)
squiana2

## The ranking matrix:
print(squiana2$rm, quote = FALSE)

## Third order habitat selection
## (relocation within home range)
## We remove the second column
## (as in the paper of Aebischer et al.)
squiana3 <- compana(squirrel$locs[,-2], squirrel$mcp[,-2])
squiana3

## The ranking matrix:
print(squiana3$rm, quote = FALSE)
```

`convnum` *Conversion from Factor to Numeric for Raster Map*

Description

The objects of class `kasc` may contain maps of type `"numeric"` (e.g. the elevation) or of type `"factor"` (e.g. the type of vegetation). `convnum` centres and scales the maps of type `"numeric"`, and factor maps are transformed into a collection of k maps of type `"numeric"` (where k is the number of levels of the factor), with zero means and unit variance (see `dudi.mix` in the `ade4` package for further details).

Usage

```
convnum(kasc)
```

Arguments

`kasc`	an object of class `kasc`

Value

Returns a list with the following components:

`kasc`	an object of class `kasc`
`weight`	the weights associated with each map of the object kasc, so that the collection of maps defining a categorical variable has the same weight that a numeric variable (see examples).

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`dudi.mix`, `as.kasc`

Examples

```
## loads the data
data(puechabon)
kasc <- puechabon$kasc
image(kasc)

## scales all the variables
toto <- convnum(kasc)
image(toto$kasc)

## the Aspect has four levels:
## four variables have been defined.
toto$weight

## The sum of the weights given to each
## level of aspect is equal to:
toto$weight[2:5]
sum(toto$weight[2:5])

## The same weight is therefore given to the factor variable Aspect and
## to any continuous variable, e.g. the elevation.
```

`count.points.id` *Number of Points in Each Pixel of a Raster Map*

Description

`count.points` counts the number of points in each pixel of a raster map of class `kasc` or `asc`. `count.points.id` counts the number of points falling in each pixel of a raster map of class `kasc` or `asc`, for different sets of points (e.g. the relocations of several animals monitored using radio-tracking)

Usage

```
count.points(xy, w)
count.points.id(xy, id, w)
```

Arguments

`xy`	a data frame containing the x and y coordinates of the points.
`id`	a factor giving, for each point, the membership of a point to a set.
`w`	an object of class `asc`, `kasc` or `mapattr`.

Value

`count.points` returns an object of class `asc` containing the number of points in each cell of the raster map.
`count.points.id` returns an object of class `kasc`, with one column per level of the factor `id`, containing the number of points numbered in each cell of the raster map.

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`as.kasc` for additionnal information on objects of class `kasc`, and `storemapattr` for further information on objects of class `mapattr`.

Examples

```
data(puechabon)
kasc <- puechabon$kasc
locs <-  puechabon$locs

## Counts the number of relocations of each wild boar
## per pixel of the raster map
(nlocrast <- count.points.id(locs[,4:5], locs[,1], kasc))
image(nlocrast)

## Counts the number of all relocations
## per pixel of the raster map
(nlocrast <- count.points(locs[,c("X","Y")], kasc))
image(nlocrast)
```

`distfacmap` *Compute distances to the different levels of a factor map*

Description

This function computes map of distances to patches belonging to the different levels of a map of class `asc` and of type `factor`.

Usage

```
distfacmap(x)
```

Arguments

`x`	an object of class `asc` and of type `factor`

Value

An object of class `kasc`.

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`import.asc` for further info on objects of class `asc`.

Examples

```
## Not run:
data(puechabon)
asp <- getkasc(puechabon$kasc, "Aspect")
image(asp)
sor <- distfacmap(asp)
image(sor)
## End(Not run)
```

`domain` *Estimation of the Potential Distribution of a Species*

Description

`domain` uses the DOMAIN algorithm to estimate the potential distribution of a species based on a list of species occurrences and on maps of the area.

Usage

```
domain(kasc, pts, type = c("value", "potential"), thresh = 0.95)
```

Arguments

`kasc`	an object of class `kasc`
`pts`	a data frame giving the x and y coordinates of the species occurrences.
`type`	a character string. The `"value"` of the suitability may be returned or the `"potential"` area of distribution
`thresh`	if `value = "potential"`, a threshold value should be supplied for the suitability (by default 0.95)

Details

This function implements the DOMAIN algorithm described in Carpenter et al. (1993).

Value

Returns a matrix of class `asc`.

Warning

`domain` is restricted to maps containing only numerical variables (i.e. no factors).

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

References

Carpenter, G., Gillison, A.N. and Winter, J. (1993) DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodiversity and conservation*, **2**, 667–680.

See Also

`as.kasc` for additionnal information on objects of class `kasc`, `import.asc` for information on matrices of class `asc`.

Examples

```
## Preparation of the data
data(puechabon)
kasc <- puechabon$kasc
kasc$Aspect <- NULL
pts <- puechabon$locs[puechabon$locs$Name == "Brock", 4:5]

## View of the data
elevation <- getkasc(kasc, "Elevation")
image(elevation)
points(pts, col = "red", pch = 16)

## Estimation of habitat suitability map
hsm <- domain(kasc, pts)
image(hsm, col = grey((1:256)/256))
contour(hsm, add = TRUE)
## Lighter areas are the most preferred areas

## Potential distribution
hsm <- domain(kasc, pts, type = "potential")
image(elevation, main = "Habitat suitability map")
image(hsm, add = TRUE, col = "orange")
points(pts, col = "red", pch = 16)
```

`enfa` *Ecological-Niche Factor Analysis*

Description

`enfa` performs an Ecological-Niche Factor Analysis. `hist.enfa` draws histograms of the row scores or of the initial variables of the ENFA. `data2enfa` prepares data (kasc and localizations) to be analyzed by the ENFA.

Usage

```
enfa(tab, pr, scannf = TRUE, nf = 1)
hist.enfa(x, scores = TRUE, type = c("h", "l"), adjust = 1, Acol, Ucol,
          Aborder, Uborder, Alwd = 1, Ulwd = 1, ...)
data2enfa(kasc, pts)
```

Arguments

`tab`	a data frame describing the available units
`pr`	a vector giving the utilization weights associated to each unit
`scannf`	logical. Whether the eigenvalues barplot should be displayed
`nf`	an integer indicating the number of kept specialization axes
`x`	an object of class `enfa`
`scores`	logical. If `TRUE`, the histograms display the row scores of the ENFA. If `FALSE`, they display the niche on the environmental variables (in this case, this is equivalent to `histniche`)
`type`	what type of plot should be drawn. Possible types are: `"h"` for histograms, `"l"` for kernel density estimates (see `?density`). By default, `type = "h"` is used. If `type = "l"` is used, the position of the mean of each distribution is indicated by dotted lines
`adjust`	if `type = "l"`, a parameter used to control the bandwidth of the density estimates (see `?density`)
`Acol`	if `type = "h"`, a color to be used to fill the histogram of the available pixels. if `type = "l"`, a color to be used for the kernel density estimates of the available pixels
`Ucol`	if `type = "h"`, a color to be used to fill the histogram of the used pixels. if `type = "l"`, a color to be used for the kernel density estimates of the used pixels
`Aborder`	color for the border of the histograms of the available pixels
`Uborder`	color for the border of the histograms of the used pixels
`Alwd`	if `type = "l"`, the line width of the kernel density estimates of the available pixels
`Ulwd`	if `type = "l"`, the line width of the kernel density estimates of the used pixels
`kasc`	a raster map of class `kasc`
`pts`	a data frame with two columns, giving the coordinates of the species locations
`...`	further arguments passed to or from other methods

Details

The niche concept, as defined by Hutchinson (1957), considers the ecological niche of a species as an hypervolume in the multidimensional space defined by environmental variables, within which the populations of a species can persist. The Ecological-Niche Factor Analysis (ENFA) has been developped by Hirzel et al. (2002) to analyse the position of the niche in the ecological space. Nicolas Perrin (1984) described the position of the niche in the n-dimensional space using two measures: the M-specialization (hereafter termed marginality) and the S-specialization (hereafter termed specialization). The marginality represents the squared distance of the niche barycentre from the mean available habitat. A large specialization corresponds to a narrow niche relative to the habitat conditions available to the species.

The ENFA first extracts an axis of marginality (vector from the average of available habitat conditions to the average used habitat conditions). Then the analysis extracts successives orthogonal axes (i.e. uncorrelated), which maximises the specialization of the species. The calculations used in the function are described in Hirzel et al. (2002).

Value

`enfa` returns a list of class `enfa` containing the following components:

`call`	original call.
`tab`	a data frame with n rows and p columns.
`pr`	a vector of length n containing the number of points in each pixel of the map.
`nf`	the number of kept specialization axes.
`m`	the marginality (squared length of the marginality vector).
`s`	a vector with all the eigenvalues of the analysis.
`lw`	row weights, a vector with n components.
`li`	row coordinates, data frame with n rows and nf columns.
`l1`	row normed scores, data frame with n rows and nf columns.
`co`	column coordinates, data frame with p rows and nf columns.
`c1`	column normed scores, data frame with p rows and nf columns.
`mar`	coordinates of the marginality vector.
`tab`	a data frame with n rows and p columns.
`pr`	a vector of length n containing the number of points in each pixel of the map.
`index`	an integer vector giving the position of the rows of `tab` in the initial object of class `kasc`.
`attr`	an object of class `mapattr` with the attributes of the initial `kasc`.

Author(s)

Mathieu Basille ⟨basille@biomserv.univ-lyon1.fr⟩

References

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

Perrin, N. (1984) Contribution à l'écologie du genre Cepaea (Gastropoda) : Approche descriptive et expérimentale de l'habitat et de la niche écologique. Thèse de Doctorat. Université de Lausanne, Lausanne.

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

See Also

`niche`, `kselect` for other types of analysis of the niche, when several species are under studies, `niche.test` to perform a test of the marginality and the tolerance of the niche, and `scatter.enfa` to have a graphical display of objects of class `enfa`.

Examples

```
## Not run:
data(lynxjura)

map <- lynxjura$map

## We keep only "wild" indices.
tmp <- lynxjura$locs[,4]!="D"
locs <- lynxjura$locs[tmp, c("X","Y")]
hist(map, type = "l")
## The variable artif is far from symetric

## We perform a square root transformation
## of this variable
## We therefore normalize the variable 'artif'
map[,4] <- sqrt(map[,4])
hist(map, type = "l")

## We prepare the data for the ENFA
(dataenfa1 <- data2enfa(map, locs[tmp, c("X","Y")]))

## We then perform the ENFA
(enfa1 <- enfa(dataenfa1$tab, dataenfa1$pr,
               scannf = FALSE))
hist(enfa1)
hist(enfa1, scores = FALSE, type = "l")

## randomization test and scatterplot
(renfa <- randtest(enfa1))
plot(renfa)
scatter(enfa1)
## End(Not run)
```

`getXYcoords` *Computes the X and Y Coordinates of the Pixels of a Raster Map*

Description

`getXYcoords` computes the geographical coordinates of the rows and columns of pixels of a raster map.

Usage

```
getXYcoords(w)
```

Arguments

`w`	an object of class `asc`, `kasc`, `sahrlocs`, or `mapattr`.

Value

Returns a list with two components:

`x`	the x coordinates of the columns of pixels of the map
`y`	the y coordinates of the rows of pixels of the map

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`import.asc`, `as.kasc`, `as.sahrlocs`, `storemapattr`

Examples

```
data(puechabon)
(elev <- getkasc(puechabon$kasc, "Elevation"))
(coords <- getXYcoords(elev))
nrow(elev) == length(coords$x)
ncol(elev) == length(coords$y)
```

`getascattr` *Copy the Attributes of an Object of Class 'asc' or 'kasc' to another Object*

Description

`getascattr` copies the attributes of an object of class `asc` to another matrix of the same size. `getkascattr` performs the same operation for objects of class `kasc`.

Usage

```
getascattr(xfrom, xto, type = c("numeric", "factor"), lev = NULL)
getkascattr(xkfrom, xkto)
```

Arguments

`xfrom`	an object of class `asc`
`xto`	a matrix with the same number of rows and columns as `xfrom`
`type`	a character string giving the type of the map (`"factor"` for maps of categorical variables, and `"numeric"` otherwise)
`lev`	if `type = "factor"`, a character vector giving the levels of the mapped variable (see `help(import.asc)`)
`xkfrom`	an object of class `kasc`
`xkto`	a data frame with the same number of rows and columns as `xkfrom`

Value

`getascattr` returns a raster matrix of class `asc`,
`getkascattr` returns a data frame of class `kasc`

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`import.asc`, `as.kasc`

Examples

```
data(puechabon)

## my.map is a map of elevation
my.map <- getkasc(puechabon$kasc, "Elevation")
sl <- getkasc(puechabon$kasc, "Slope")
attributes(sl) <- NULL
sl <- matrix(sl, ncol = ncol(my.map))

## sl is a matrix with the same size as my.map
toto <- getascattr(my.map, sl)
image(toto)

## same rationale with aspect
asp <- getkasc(puechabon$kasc, "Aspect")
le <- levels(asp)
attributes(asp) <- NULL
asp <- matrix(asp, ncol = ncol(my.map))

## asp is now a matrix with the same size as my.map
tutu <- getascattr(my.map, asp, typ = "factor", lev = le)
cl <- colasc(tutu, NorthEast = "blue", SouthEast = "red",
           SouthWest = "orange", NorthWest = "green")
image(tutu, clfac = cl)
```

`getcontour` *Computes the Contour Polygon of a Raster Object*

Description

`getcontour` computes the contour polygon of a raster object of class `asc`. When the object is made of several parts, the function returns one polygon per part.

Usage

```
getcontour(x)
```

Arguments

`x`	an object of class `asc`

Value

Returns an object of class `area`.

Warning

Holes in the polygons are not taken into account by the function.

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`import.asc` for additionnal information on objects of class `asc`, `as.area` for information on objects of class `area`, and `area2dxf` to export the results toward a GIS.

Examples

```
data(puechabon)
kasc <- puechabon$kasc

#########################################
## Example with one object:
## Gets the first map of the "kasc" object
## Map of the elevation
elev <- getkasc(kasc, "Elevation")
image(elev)

## Get the contour polygon
cpol <- getcontour(elev)

## Draw the polygon
i <- cpol[,2:3]
polygon(i, col = "green", lwd = 2)

#########################################
## Example with two objects:
## home ranges of wild boar

hr <- getsahrlocs(puechabon$sahr, "hr")
u <- getkasc(hr, "Jean")
image(u)

## Get the contour polygons
p <- getcontour(u)
plot(p, lwd = 2)
```

`hist.kasc` *Histograms of Mapped Variables*

Description

`hist.kasc` performs histograms of the variables mapped in objects of class `kasc`.

Usage

```
hist.kasc(x, type = c("h", "l"), adjust = 1, col = "blue", ...)
```

Arguments

`x`	a raster map of class `kasc`
`type`	what type of plot should be drawn. Possible types are: `"h"` for histograms, `"l"` for kernel density estimates (see `?density`). By default, `type = "h"` is used. If `type = "l"` is used, the position of the mean of each distribution is indicated by dotted lines
`adjust`	if `type = "l"`, a parameter used to control the bandwidth of the density estimate (see `?density`)
`col`	color for the histogram
`...`	further arguments passed to or from other methods

Author(s)

Mathieu Basille ⟨basille@biomserv.univ-lyon1.fr⟩

See Also

`as.kasc`

Examples

```
## Example with factors and numeric variables
data(puechabon)
hist(puechabon$kasc)

## Removing the factor Aspect, and smoothing gives :
hist(puechabon$kasc[,-2], type = "l")
```

`histniche` *Histograms of the Ecological Niche*

Description

`histniche` draws histograms of the variables mapped in an object of class `kasc` (habitat available for the species). The histograms of the habitat used by a species (i.e. the niche) are computed from the locations of the focus species, and are added to the plot of available habitat for each variable, for comparison.

Usage

```
histniche(kasc, pts, type = c("h", "l"), adjust = 1,
          Acol, Ucol, Aborder, Uborder, Alwd = 1, Ulwd = 1, ...)
```

Arguments

`kasc`	a raster map of class `kasc`
`pts`	a data frame with two columns, giving the coordinates of the species locations
`type`	what type of plot should be drawn. Possible types are: `"h"` for histograms, `"l"` for kernel density estimates (see `?density`). By default, `type = "h"` is used. If `type = "l"` is used, the position of the mean of each distribution is indicated by dotted lines
`adjust`	if `type = "l"`, a parameter used to control the bandwidth of the density estimate (see `?density`)
`Acol`	color for the histograms of the available pixels
`Ucol`	color for the histograms of the used pixels
`Aborder`	if `type = "h"`, color for the border of the histograms of the available pixels (see `help(hist.default)`)
`Uborder`	if `type = "h"`, color for the border of the histograms of the used pixels (see `help(hist.default)`)
`Alwd`	if `type = "l"`, line width for the density estimate of the available pixels
`Ulwd`	if `type = "l"`, line width for the density estimate of the used pixels
`...`	further arguments passed to or from other methods

Author(s)

Mathieu Basille ⟨basille@biomserv.univ-lyon1.fr⟩

Examples

```
## Not run:
## Example with factors and numeric variables
data(puechabon)
histniche(puechabon$kasc, puechabon$locs[, c("X", "Y")])

## Aspect is a factor, then it's not possible to use
## kernel density estimates for it :
histniche(puechabon$kasc, puechabon$locs[, c("X", "Y")], type = "l")
## End(Not run)
```

`hr.rast` *Rasterisation of Objects of Class 'area'*

Description

Converts an object of class `area` (used in many functions of the package `ade4`) to an object of class `kasc` (rasterisation).

Usage

```
hr.rast(mcp, w)
```

Arguments

`mcp`	an object of class `area`
`w`	a raster map of class `kasc` or of class `asc`

Value

Returns an object of class `kasc`.

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`as.kasc` for additional information on objects of class `kasc`, `as.area` for further information on the class `area`

Examples

```
data(puechabon)
kasc <- puechabon$kasc
locs <-  puechabon$locs

## Computes the home range of the animals
cp <- mcp(locs[,4:5], locs[,1])
area.plot(cp)

## Converts the home range to raster
cprast <- hr.rast(cp, kasc)
image(cprast)
```

`image.asc` *Displays a Color Image of an Object of Class 'asc'*

Description

These functions display a raster matrix of class `asc`.

Usage

```
image.asc(x, clfac = NULL, col = gray((240:1)/256), ...)
contour.asc(x, ...)
persp.asc(x, ...)
plot.asc(x, ...)
```

Arguments

`x`	an object of class `asc`
`clfac`	for maps of type `"factor"`, a character vector giving the names of colors for each level of the factor (see `help(colasc)`)
`col`	for maps of type `"numeric"`, the colors to be used (see `help(par)`)
`...`	additional arguments to be passed to the generic function `image`, `persp`, `contour`, and `filled.contour` (but see below)

Note

The function `plot.asc` uses the function `filled.contour`. The output produced by `filled.contour` is actually a combination of two plots; one is the filled contour and one is the legend. Two separate coordinate systems are set up for these two plots, but they are only used internally - once the function has returned these coordinate systems are lost. If you want to annotate the main contour plot, for example to add points, you can specify graphics commands in the `plot.axes` argument of the function `filled.contour` (this argument is to be passed to the function `plot.asc`). An example is given below.

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`image`, `contour`, `persp`, `filled.contour`, `import.asc`

Examples

```
data(puechabon)

# Case of a continuous variable: the elevation
my.map <- getkasc(puechabon$kasc, "Elevation")
image(my.map, main = "Elevation in Puechabon")
contour(my.map, add = TRUE)

# use of the function plot.asc: adding points on the map
plot(my.map, plot.axes = {points(puechabon$locs[,c("X","Y")])},
     main = "Elevation")

# Case of a factor: the aspect
asp <- getkasc(puechabon$kasc, "Aspect")
cl <- colasc(asp, NorthEast = "blue", SouthEast = "red",
             SouthWest = "orange", NorthWest = "green")

## graphical display
image(asp, clfac = cl, main = "Aspect", xlab = "Lambert X",
      ylab = "Lambert Y")
legend(706500, 3162000, legend = levels(asp), fill = cl,
       cex = 0.7)

opar<-par(mar = c(0,0,3,0), bg = "slategray")
persp(my.map, scale = FALSE, box = FALSE, border = NA, shade = 0.75,
      col = "darkolivegreen3", expand = 2, theta = -60, phi = 30,
      main = "The topography of Puechabon")
par(opar)
```

`image.sahrlocs` *Graphical Display of the Habitat Composition of the Home Ranges of Animals Monitored Using Radio-Tracking*

Description

`image.sahrlocs` allows a gray-level display of the composition of home ranges (different colors are used for factors). For a given variable, the minimum gray level (default is `"white"`) and the maximum gray level (default is `"black"`) represents respectively the minimum and the maximum of the variable ****on the study area****.

Usage

```
image.sahrlocs(x, ani = names(x$hr), var = names(x$sa),
               mar = c(0, 0, 0, 0), axes = FALSE, dfidxy = NULL, colpts =
               "black", pch = 21, bg = "white", inv = FALSE, cexpts = 0.6,
               csub = 2, possub = c("bottomleft", "bottomright", "topleft",
               "topright"), ...)
```

Arguments

`x`	an object of class `sahrlocs`
`ani`	a character vector giving the names of the variables of the `"hr"` component (the animals) for which a display is wanted
`var`	a character vector giving the names of the variables of the `"sa"` component (the habitat variables) for which a display is wanted
`mar`	the graphical parameter `mar` (see `par`)
`axes`	logical. Whether the axes should be plotted
`csub`	the character size for the legend, used with `par("cex")*csub`
`possub`	a character string indicating the sub-title position (`"topleft"`, `"topright"`, `"bottomleft"`, `"bottomright"`)
`dfidxy`	an optional data frame with three columns giving the identity and the coordinates of the relocations of each animal. (if not `NULL`, the relocations of each animal are plotted in its home range, see Examples)
`colpts`	if `dfidxy` is not `NULL`, the color of the points to be used for the plot of the relocations
`pch`	if `dfidxy` is not `NULL`, the size of the points to be used for the plot of the relocations (see `par`)
`bg`	if `dfidxy` is not `NULL`, the background color to be used for the plot of the relocations (see `par`)
`inv`	by default, lower values of the mapped variables are brighter. If `FALSE`, the lower values are darker
`cexpts`	if `dfidxy` is not `NULL`, the size of the points
`...`	additionnal parameters to be passed to the generic function `image`

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`as.sahrlocs` for additionnal information on objects of class `sahrlocs`

Examples

```
data(puechabon)
sahr <- puechabon$sahr
sahr

## Displays all the variables for a given animal
image(sahr, ani = "Chou")

## Displays all the animals for a given variable
image(sahr, var = "Elevation")

## Load and displays the relocations of the animals
locs <-  puechabon$locs[,c(1,4:5)]
image(sahr, var = "Elevation", dfidxy = locs, pch = 21)
```

`import.asc` *Arcview ASCII Raster File Importation And Exportation*

Description

`import.asc` imports Arcview ASCII raster file created with the Arcview 3.2 command `"File -> Export data source"`, or the Arcview 8 command `"Arc Toolbox -> Conversion tools -> From raster -> Raster to ASCII"`.
Conversely, `export.asc` allows to export an `asc` object toward Arcview.
ASCII raster files may also be created using the free program landserf (`http://www.soi.city.ac.uk/~jwo/landserf/`).

Usage

```
import.asc(file, type = c("numeric", "factor"), lev = NULL,
           levnb = 1, labnb = 3)
export.asc(x, file)
as.asc(x, xll = 1, yll = 1, cellsize = 1, type = c("numeric", "factor"),
       lev = levels(factor(x)))
print.asc(x, ...)
```

Arguments

`file`	a character string. The name of an Arcview ASCII Raster file
`type`	a character string. Either `"numeric"` or `"factor"`
`lev`	if `type = "factor"`, either a vector giving the labels of the factor levels, or the name of a file giving the correspondence table of the map (see details)
`levnb`	if `lev` is the name of a file containing a correspondence table exported from Arcview, the column number in this table where the factor levels are stored (i.e. the numbers indicating the levels of the factor)
`labnb`	if `lev` is the name of a file containing a correspondence table exported from Arcview, the column number in this table where the factor labels are stored (i.e. the character strings indicating the labels associated with each level of the factor)
`x`	an object of class `asc`. For the function `as.asc`, a matrix
`xll`	the x coordinate of the center of the lower left pixel of the map

`yll`	the y coordinate of the center of the lower left pixel of the map
`cellsize`	the size of a pixel on the studied map
`...`	additionnal arguments to be passed to the function `print`

Details

The raster maps may be of type `"numeric"` (e.g. elevation on an area) or `"factor"` (e.g. the type of vegetation on an area). If the map is of type `factor`, the levels should be indicated. The `".asc"` files store the values of the mapped variable with numeric values. Each level of the factor is coded on the map by a number. The argument `lev` of `import.asc` or `as.asc` gives the labels corresponding to each number. Alternatively, these levels may be specified using a correspondence table exported from Arcview (with this software, command `"Theme -> table"`, then `"File -> Export"`, and finally export in delimited text format). An example of such file is provided in the directory "ascfiles" of the package, see the examples below. `export.asc` allows only exportation of numeric maps.

Value

Returns a raster matrix of the class `asc`, with the following attributes :

`xll`	the x coordinate of the center of the lower left pixel of the map
`yll`	the y coordinate of the center of the lower left pixel of the map
`cellsize`	the size of a pixel on the studied map
`type`	either `"numeric"` or `"factor"`.
`levels`	if `type = "factor"`, the levels of the factor.

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

References

Arcview: `http://www.esri.com`
Landserf: `http://www.soi.city.ac.uk/~jwo/landserf/`

See Also

`image.asc`

Examples

```
## Not run:
## Importation of asc files: numeric
## Path of the file to be imported
(file1 <-  paste(system.file(package = "adehabitat"),
               "ascfiles/Elevation.asc", sep = "/"))

el <- import.asc(file1)
image(el)
el

## Importation of asc files: factor
(file2 <- paste(system.file(package = "adehabitat"),
```

```
                 "ascfiles/Aspect.asc", sep = "/"))
(levelfile <- paste(system.file(package = "adehabitat"),
                 "ascfiles/Aspect.txt", sep = "/"))
asp <- import.asc(file2, lev = levelfile, type = "factor")
image(asp)
asp

## map of white noise
wafwaf <- matrix(rnorm(10000), 100, 100)
wafwaf <- as.asc(wafwaf)
image(wafwaf)

## exportation of a map
export.asc(wafwaf, "foo.asc")

## remove the created file:
file.remove("foo.asc")
## End(Not run)
```

`join.asc` *Finds the Value of Mapped Variables at some Specified Locations (Spatial Join)*

Description

`join.asc` finds the value of a mapped variable at some specified locations.
`join.kasc` is the same function as `join.asc`, with several maps.

Usage

```
join.asc(pts, x)
join.kasc(pts, w)
```

Arguments

`x`	an object of class `asc`
`w`	an object of class `kasc`
`pts`	the x and y coordinates of points to be placed on the map.

Value

`join.asc` returns a vector with length equals to the number of points in `pts`.
`join.kasc` returns a data frame with a number of columns equals to the number of variables in the object of class `kasc`, and with each row corresponding to the rows of `pts`.

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`as.kasc` and `import.asc`

Examples

```
data(puechabon)
x <- puechabon$kasc

## for each relocation, finds the values of the variables in x
toto <- join.kasc(pts = cbind(puechabon$locs$X, puechabon$locs$Y), x)
toto[1:4,]
```

`kasc2df` *Conversion of Objects of Class kasc*

Description

An object of class `kasc` stores several maps in a data frame (one column per variable, and one row per pixel of the raster map). However, the features mapped are rarely rectangle-shaped, whereas the map are inevitably rectangles. Therefore, a lot of pixels of the maps do not contain data. The pixels of the map that do not contain data are `NA` in this data frame. `kasc2df` will "clean" the object of class `kasc` from these NAs, and will return a data frame containing only mapped values that can be used in subsequent analysis.
After these analyses, `df2kasc` may be used to convert the modified data frame to an object of class `kasc` for mapping (e.g. for maps of factorial axes, using `dudi` analyses, see `help(dudi)`).

Usage

```
kasc2df(x, var = names(x))
df2kasc(df, index, x)
```

Arguments

`x`	an object of class `kasc` in `kasc2df` an object of class `kasc` or `mapattr` in `df2kasc`
`var`	a character vector. The names of the variables in the `kasc` that are to be kept in the output
`df`	a data frame resulting of a computation of the component `tab` of the list previously returned by the `kasc2df` (see section `Value` below). This computation may be any form of analysis (Principal component analysis, modelling techniques, etc.)
`index`	an integer vector giving the position of the rows of `df` in the returned `kasc` (such an index can be computed using `kasc2df`)

Value

`kasc2df` returns a list with the following components:

`tab`	a data frame without NAs, with a number of variables equals to `length(var)`.
`index`	a vector of indices of the rows of the `kasc` kept for the analyses (that is, not `NA`).

`df2kasc` returns an object of class `kasc`.

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`as.kasc` for additional information on objects of class `kasc`.

Examples

```
data(puechabon)
kasc <- puechabon$kasc

# Display the kasc object
image(kasc)

# Preparation for Principal component analysis
x <- kasc2df(kasc)
x$tab <- x$tab[, (names(x$tab) != "Aspect")]

# Principal component analysis
ana <- dudi.pca(x$tab, scannf = FALSE)

s.arrow(ana$co)
s.label(ana$li, clab = 0)

## Map of the scores of the rows
scores <- df2kasc(ana$li, x$index, kasc)
image(scores)
```

`kasc2spixdf` *Conversion of maps from/to the package "sp"*

Description

These functions convert maps of classes available in adehabitat toward classes available in the package `sp` and conversely.

`kasc2spixdf` converts an object of class `kasc` into an object of class `SpatialPixelsDataFrame`.

`asc2spixdf` converts an object of class `asc` into an object of class `SpatialGridDataFrame`.

`spixdf2kasc` converts an object of class `SpatialPixelsDataFrame` or `SpatialGridDataFrame` into an object of class `asc` or `kasc`.

`area2sr` converts an object of class `area` into an object of class `SpatialRings`.

`sr2area` converts an object of class `SpatialRings` or `SpatialRingsDataFrame` into anobject of class `area`.

`attsr2area` gets the data attribute of an object of class `SpatialRingsDataFrame` and stores is into a data frame.

`traj2spdf` converts an object of class `traj` into an object of class `SpatialPointsDataFrame`.

`traj2sldf` converts an object of class `traj` into an object of class `SpatialLinesDataFrame`.

Usage

```
kasc2spixdf(ka)
asc2spixdf(a)
spixdf2kasc(sg)
area2sr(ar)
sr2area(sr)
attsr2area(srdf)
traj2spdf(tr)
traj2sldf(tr, byid = FALSE)
```

Arguments

`ka`	an object of class `kasc`.
`a`	an object of class `asc`.
`sg`	an object of class `SpatialPixelsDataFrame` or `SpatialGridDataFrame`.
`ar`	an object of class `area`.
`sr`	an object of class `SpatialRings` or `SpatialRingsDataFrame`.
`srdf`	an object of class `SpatialRingsDataFrame`.
`tr`	an object of class `traj`.
`byid`	logical. If `TRUE`, one objects of class `Slines` correspond to one animal. if `FALSE`, one object of class `Slines` correspond to one burst.

Details

We describe here more in detail the functions `sr2area` and `attsr2area`. Objects of class `area` do not deal with holes in the polygons, whereas the objects of class `SpatialRings` do. Therefore, when holes are present in the `SpatialRings` object passed as argument, the function `sr2area` ignore them and returns only the external contour of the polygon (though a warning is returned, see example).

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`import.asc` for information on objects of class `asc`, `as.kasc` for info on objects of class `kasc`, `as.area` for info on objects of class `area`, `as.traj` for objects of class `traj`.

Examples

```
## Not run:
if (require(sp) {

###########################################
##
```

```
## Conversion kasc -> SpatialPixelsDataFrame
##

data(puechabon)
toto <- kasc2spixdf(puechabon$kasc)
image(toto)
summary(toto)

#### and conversely
toto <- spixdf2kasc(toto)
image(toto)
hist(toto)

data(meuse.grid)
m <- SpatialPixelsDataFrame(points = meuse.grid[c("x", "y")],
                           data = meuse.grid)
i <- spixdf2kasc(m)
image(i)

### conversion asc -> SpatialPixelsDataFrame
cuicui <- asc2spixdf(getkasc(toto,1))
image(cuicui)

#########################################
##
## Conversion area -> SpatialRings
##

data(elec88)
ar <- as.area(elec88$area)
plot(ar)
toto <- area2sr(ar)
plot(toto)

#########################################
##
## Conversion SpatialRings -> area
##

## First create an object of class "SpatialRingsDataFrame"
data(ncshp)
nc1 <- as.SpatialRings.Shapes(nc.shp$Shapes, as.character(nc.shp$att.data$FIPS))
plotSpatialRings(nc1)
df <- nc.shp$att.data
rownames(df) <- as.character(nc.shp$att.data$FIPS)
ncSRDF <- SpatialRingsDataFrame(nc1, df)

## and then conversion:
coincoin <- sr2area(ncSRDF)
## please note the warnings

plot(coincoin)
## gets the attributes
haha <- attsr2area(ncSRDF)
area.plot(coincoin, values = df$SID74/df$BIR74)

#########################################
```

```
##
## Conversion traj -> SpatialPointsDataFrame
##

data(puechcirc)
plot(puechcirc)

toto <- traj2spdf(puechcirc)
plot(toto)

#########################################
##
## Conversion traj -> SpatialLinesDataFrame
##

data(puechcirc)
plot(puechcirc)

toto <- traj2sldf(puechcirc)
plot(toto)

}
## End(Not run)
```

`kernelUD` *Estimation of Kernel Home-Range*

Description

`kernelUD` is used to estimate the utilization distribution (UD) of animals monitored by radio-tracking, with the classical kernel method.
`getvolumeUD` and `kernel.area` provide utilities for home-range size estimation.
`getverticeshr` stores the home range contour as objects of class `area` in a list of class `kver`, with one component per animal.
`plot.kver` is used to display these home-range contours.
`kernelbb` is used to fit an utilization distribution using the brownian bridge approach of the kernel method (for autocorrelated relocations; Bullard, 1991).

Usage

```
kernelUD(xy, id = NULL, h = "href", grid = 40, same4all = FALSE,
         hlim = c(0.1, 1.5), kern = "bivnorm")
print.khr(x, ...)
image.khr(x, axes = FALSE, mar = c(0,0,2,0),
          addcontour = TRUE, addpoints = TRUE, ...)
plotLSCV(x)
getvolumeUD(x)
kernel.area(xy, id, h = "href", grid=40,
            same4all = FALSE, hlim = c(0.1,1.5), kern = "bivnorm",
            levels = seq(20,95, by = 5),
            unin = c("m", "km"),
            unout = c("ha", "km2", "m2"))
getverticeshr(x, lev = 95)
```

```
plot.kver(x, which = names(x), colpol=rainbow(length(which)),
          colborder=rep("black", length(which)), lwd = 2,
          add=FALSE, ...)
kernelbb(tr, sig1, sig2, grid = 40, same4all=FALSE, byburst=FALSE)
```

Arguments

`xy`	a data frame with two columns (x and y coordinates of the animal relocations)
`id`	an optional factor giving the animals identity associated to `xy`
`h`	a character string or a number. If `h` is set to `"href"`, the ad hoc method is used for the smoothing parameter (see details). If `h` is set to `"LSCV"`, the least-square cross validation method is used. Note that `"LSCV"` is not available if `kern = "epa"`. Alternatively, `h` may be set to any given numeric value
`grid`	a number giving the size of the grid on which the UD should be estimated. Alternatively, this parameter may be an object of class `asc` (see examples)
`same4all`	logical. If `TRUE`, the same grid is used for all animals. If `FALSE`, one grid per animal is used
`hlim`	a numeric vector of length two. If `h = "LSCV"`, the function minimizes the cross-validation criterion for values of h ranging from `hlim[1]*href` to `hlim[2]*href`, where `href` is the smoothing parameter computed with the ad hoc method (see below)
`kern`	a character string. If `"bivnorm"`, a bivariate normal kernel is used. If `"epa"`, an Epanechnikov kernel is used.
`x`	an object of class `khr` returned by `kernelUD`. For `plot.kver`, an object of class `kver` returned by `getverticeshr`
`which`	a vector of character indicating the polygons to be plotted
`colpol`	a vector of the color for filling the polygon. The default, `NA`, is to leave polygons unfilled
`colborder`	a vector of the color to draw the border. The default. Use `border = NA` to omit borders
`lwd`	the border width, a **positive** number
`add`	logical. if `TRUE`, the polygons are added to a previous plot
`axes`	logical. Whether the axes are to be plotted
`mar`	the margin parameter (see `help(par)`)
`addcontour`	logical. if `TRUE`, contours are drawn on the graphics
`addpoints`	logical. if `TRUE`, the animal relocations are drawn on the graphics
`levels`	a vector of percentage levels for home-range size estimation
`unin`	the units of the relocations coordinates. Either `"m"` for meters (default) or `"km"` for kilometers
`unout`	the units of the output areas. Either `"m2"` for square meters, `"km2"` for square kilometers or `"ha"` for hectares (default)
`lev`	the percentage level for home range contour estimation.
`tr`	an object of class `traj`.
`sig1`	first smoothing parameter for the brownian bridge method (related to the imprecision of the relocations).

`sig2`	second smoothing parameter for the brownian bridge method (related to the speed of the animals).
`byburst`	logical. Whether the brownian bridge estimation should be done by burst.
`...`	additionnal parameters to be passed to the generic functions `print` and `image`

Details

The Utilization Distribution (UD) is the bivariate function giving the probability density that an animal is found at a point according to its geographical coordinates. Using this model, one can define the home range as the minimum area in which an animal has some specified probability of being located. The functions used here correspond to the approach described in Worton (1995).

The kernel method has been recommended by many authors for the estimation of the utilization distribution (e.g. Worton, 1989, 1995). The default method for the estimation of the smoothing parameter is the *ad hoc* method, i.e. for a bivariate normal kernel

$$h = \sigma n^{-\frac{1}{6}}$$

where

$$\sigma = 0.5(\sigma(x) + \sigma(y))$$

which supposes that the UD is bivariate normal. If an Epanechnikov kernel is used, this value is multiplied by 1.77 (Silverman, 1986, p. 86). Alternatively, the smoothing parameter h may be computed by Least Square Cross Validation (LSCV). The estimated value then minimizes the Mean Integrated Square Error (MISE), i.e. the difference in volume between the true UD and the estimated UD. Note that the cross-validation criterion cannot be minimized in some cases. According to Seaman and Powell (1998) *"This is a difficult problem that has not been worked out by statistical theoreticians, so no definitive response is available at this time"* (see Seaman and Powell, 1998 for further details and tricky solutions). `plotLSCV` allows to have a diagnostic of the success of minimization of the cross validation criterion (i.e. to know whether the minimum of the CV criterion occurs within the scanned range). Finally, the UD is then estimated over a grid.

The default kernel is the bivariate normal kernel, but the Epanechnikov kernel, which requires less computer time is also available for the estimation of the UD.

The function `getvolumehr` modifies the UD component of the object passed as argument, so that the contour of the UD displayed by the functions `contour` and `image.khr` corresponds to the different percentage levels of home-range estimation (see examples). In addition, this function is used in the function `kernel.area`, to compute the home-range size. Note, that the function `plot.hrsize` (see the help page of this function) can be used to display the home-range size estimated at various levels.

The function `kernelbb` uses the brownian bridge approach to estimate the UD with autocorrelated relocations (Bullard, 1991). Instead of simply smoothing the relocation pattern, it takes into account the fact that between two successive relocations r1 and r2, the animal has moved through a connected path, which is not necessarily linear. A brownian bridge estimates the density of probability that this path passed through any point of the study area, given that the animal was located at the point r1 at time t1 and at the point r2 at time t2, with a certain amount of inaccuracy (see Examples). Brownian bridges are placed over the different sections of the traject, and these functions are then summed over the area. The brownian bridge approach therefore smoothes a trajectory.

The brownian bridge estimation relies on two smoothing parameters, `sig1` and `sig2`. The parameter `sig1` is related to the speed of the animal, and describes how far from the line joining two successive relocations the animal can go during one time unit (here the time is measured in second). The larger this parameter is, and the more wiggly the trajectory is likely to be. The parameter `sig2` is equivalent to the parameter `h` of the classical kernel method: it is related to the inaccuracy of the relocations (See examples for an illustration of the smoothing parameters).

Value

The class `khr` is a class grouping three sub-classes, `khrud`, `kbbhrud` and `khrudvol`: `kernelUD` returns a list of the class `khrud`. This list has one component per animal (named as the levels of argument `id`). Each component is itself a list, with the following sub-components:

`UD`	an object of class `asc`, with the values of density probability in each cell of the grid
`h`	if `LSCV` is not used, the value of the smoothing parameter. if `LSCV` is used, a list with three components: **`CV`** the results of the cross-validation procedure. The first column contains the sequence of values tested for the smoothing parameter, and the second column contains the value of the cross-validation criterion. **`convergence`** TRUE if the LSCV succeeds (i.e. if the optimum smoothing parameter have been found by the procedure), FALSE otherwise. **`h`** the value of the smoothing parameter used in UD estimation.
`locs`	The relocations used in the estimation procedure.
`hmeth`	The argument `h` of the function kernelUD

`kernelbb` returns an object of class `kbbhrud`, with the same components as lists of class `khrud`.

`getvolumeUD` returns a list of class `khrvol`, with the same components as lists of class `khrud`.

`kernel.area` returns a data frame of subclass `"hrsize"`, with one column per animal and one row per level of estimation of the home range.

`getverticeshr` returns an object of class `kver`.

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

References

Bullard, F. (1991) *Estimating the home range of an animal: a Brownian bridge approach.* Master of Science, University of North Carolina, Chapel Hill.

Silverman, B.W. (1986) *Density estimation for statistics and data analysis*. London: Chapman & Hall.

Worton, B.J. (1989) Kernel methods for estimating the utilization dirstibution in home-range studies. *Ecology*, **70**, 164–168.

Worton, B.J. (1995) Using Monte Carlo simulation to evaluate kernel-based home range estimators. *Journal of Wildlife Management*, **59**,794–800.

Seaman, D.E. and Powell, R.A. (1998) *Kernel home range estimation program (kernelhr)*. Documentation of the program. `ftp://ftp.im.nbs.gov/pub/software/CSE/wsb2695/KERNELHR.ZIP`.

See Also

`import.asc` for additionnal informations on objects of class `asc`, `mcp` for estimation of home ranges using the minimum convex polygon, and for help on the function `plot.hrsize`.

Examples

```
data(puechabon)
loc <- puechabon$locs[, c("X", "Y")]
id <- puechabon$locs[, "Name"]

## Estimation of UD for the four animals
ud <- kernelUD(loc, id)
ud

image(ud) ## Note that the contours
          ## corresponds to values of probability density
udvol <- getvolumeUD(ud)
image(udvol)
## Here, the contour corresponds to the
## home ranges estimated at different probability
## levels (i.e. the contour 90 corresponds to the 90 percent
## kernel home-range)
## udvol describes, for each cell of the grid,
## the smaller home-range to which it belongs

## Calculation of the 95 percent home range
ver <- getverticeshr(ud, 95)
elev <- getkasc(puechabon$kasc, "Elevation") # Map of the area
image(elev)
plot(ver, add=TRUE)
legend(696500, 3166000, legend = names(ver), fill = rainbow(4))

## Example of estimation using LSCV
udbis <- kernelUD(loc, id, h = "LSCV")
image(udbis)

## Compare the estimation with ad hoc and LSCV method
## for the smoothing parameter
(cuicui1 <- kernel.area(loc, id)) ## ad hoc
plot(cuicui1)
(cuicui2 <- kernel.area(loc, id, h = "LSCV")) ## LSCV
plot(cuicui2)

## Diagnostic of the cross-validation
plotLSCV(udbis)

## Use of the same4all argument: the same grid
## is used for all animals
udbis <- kernelUD(loc, id, same4all = TRUE)
image(udbis)

## And finally, estimation of the UD on a map
## (e.g. for subsequent analyses on habitat selection)
elev <- getkasc(puechabon$kasc, "Elevation")
opar <- par(mfrow = c(2, 2), mar = c(0, 0, 2, 0))
```

```
cont <- getcontour(elev)

for (i in 1:length(udbis)) {
   image(elev, main = names(udbis)[i], axes = FALSE)
   points(udbis[[i]]$locs, pch = 21, bg = "white", col = "black")
}

## Measures the UD in each pixel of the map
udbis <- kernelUD(loc, id, grid = elev)
opar <- par(mfrow = c(2, 2), mar = c(0, 0, 2, 0))
for (i in 1:length(udbis)) {
  image(udbis[[i]]$UD, main = names(udbis)[i], axes = FALSE)
  box()
  polygon(cont[, 2:3])
}
par(opar)



###############################################
###############################################
###############################################
###
###         Kernel estimation: a brownian
###               bridge approach

## Not run:

## loads the data
data(puechcirc)

## gets one circuit
x <- getburst(puechcirc, burst = "CH930824")

## fits the home range
(tata <- kernelbb(x, 10, 10, same4all = TRUE))
image(tata)
lines(x$x, x$y, lwd = 2, col = "red")

## Image of a brownian bridge. Fit with two relocations:
fac <- factor(c("a", "a"))
xx <- c(0,1)
yy <- c(0,1)
date <- c(0,1)
class(date) <- c("POSIXt", "POSIXct")
tr <- as.traj(fac, data.frame(x = xx,y = yy), date)

## Use of different smoothing parameters
sig1 <- c(0.05, 0.1, 0.2, 0.4, 0.6)
sig2 <- c(0.05, 0.1, 0.2, 0.5, 0.7)

y <- list()
for (i in 1:5) {
  for (j in 1:5) {
     k <- paste("s1=", sig1[i], ", s2=", sig2[j], sep = "")
     y[[k]]<-kernelbb(tr, sig1[i], sig2[j])[[1]]$UD
   }
```

```
 }

## Displays the results
opar <- par(mar = c(0,0,2,0), mfrow = c(5,5))
foo <- function(x)
   {
     image(y[[x]], main = names(y)[x], axes = F)
     points(tr[,c("x","y")], pch = 16, col = "red")
   }
lapply(1:length(y), foo)

par(opar)

## End(Not run)
```

`kselect` *K-Select Analysis: a Method to Analyse the Habitat Selection by Animals*

Description

Performs a multivariate analysis of ecological data (K-select analysis).

Usage

```
kselect(dudi, factor, weight, scannf = TRUE, nf = 2, ewa = FALSE)
print.kselect(x, ...)
kplot.kselect(object, xax = 1, yax = 2, csub = 2, possub = c("topleft",
               "bottomleft", "bottomright", "topright"),
               addval = TRUE, cpoint = 1, csize = 1, clegend = 2, ...)
hist.kselect(x, xax = 1, mar = c(0, 0, 0, 0), ampl = 1,
             col.out = gray(0.75), col.in = gray(0.75), ncell = TRUE,
             denout = NULL, denin = NULL, lwdout = 1, lwdin = 1,
             maxy = 1, csub = 2, possub =
             c("bottomleft", "topleft", "bottomright", "topright"),
             ncla = 15, ...)
plot.kselect(x, xax = 1, yax = 2, ...)
```

Arguments

`dudi`	an object of class `dudi`
`factor`	a factor with the same length as `nrow(dudi$tab)`
`weight`	a numeric vector of integer values giving the weight associated to the rows of `dudi$tab`.
`scannf`	logical. Whether the eigenvalues bar plot should be displayed
`nf`	if `scannf = FALSE`, an integer indicating the number of kept axes
`ewa`	logical. If `TRUE`, uniform weights are given to all animals in the analysis. If `FALSE`, animal weights are given by the proportion of relocations of each animal (i.e. an animal with 10 relocations has a weight 10 times lower than an animal with 100 relocations)
`x`	an object of class `kselect`

`object`	an object of class `kselect`
`xax`	the column number for the x-axis
`yax`	the column number for the y-axis
`addval`	logical. If `TRUE`, the frequency of the relocations per animal is displayed (see examples)
`cpoint`	the size of the points (if 0, the points where no relocations are found are not displayed)
`mar`	the margin parameter (see `help(par)`).
`ampl`	the amplification factor (i.e. `ylim = c(-1 , 1) / ampl`)
`col.out`	character string. The color of the upper histogram
`col.in`	character string. The color of the lower histogram
`ncell`	logical. If `TRUE`, the histogram shows the distribution of the cells of the raster map where at least one relocation is found. If `FALSE`, the histogram shows the distribution of the relocations
`denout`	the density of shading lines for the upper histogram, in lines per inch (see `help(hist)` for further informations)
`denin`	the density of shading lines for the lower histogram, in lines per inch
`lwdout`	the line width for the upper histogram
`lwdin`	the line width for the lower histogram
`maxy`	the maximum Y coordinate (since the histogram draws frequencies, default value of `maxy` is 1)
`csub`	the character size for the legend, used with `par("cex")*csub`
`csize`	the size coefficient for the points
`clegend`	the character size for the legend used by `par("cex")*clegend`
`possub`	a character string indicating the sub-title position (`"topleft", "topright", "bottomleft", "bottomright"`)
`ncla`	the number of classes of the histograms
`...`	additional arguments to be passed to the generic function `hist`, `print` or, in the case of `plot.kselect`, `s.distri`

Value

`kselect` returns a list of the class `kselect` and `dudi` (see `dudi`).

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

References

Calenge, C., Dufour, A.B. and Maillard, D. (*submitted*). K-select analysis, a new method to analyse habitat selection in radio-tracking studies.

See Also

`sahrlocs2kselect` for conversion of objects class `sahrlocs` to objects suitable for a K-select analysis, `s.distri`, and `dudi` for class `dudi`.

Examples

```
## Not run:
## Loads the data
data(puechabon)
sahr <- puechabon$sahr

## prepares the data for the kselect analysis
x <- sahrlocs2kselect(sahr)
tab <- x$tab

## Example of analysis with two variables: the slope and the elevation.
## Have a look at the use and availability of the two variables
## for the 4 animals
tab <- tab[,((names(tab) == "Slope")|(names(tab) == "Elevation"))]
tab <- scale(tab)
tmp <- split.data.frame(tab, x$factor)
wg <- split(x$weight, x$factor)
opar <- par(mfrow = n2mfrow(nlevels(x$factor)))
for (i in names(tmp))
  s.distri(scale(tmp[[i]]), wg[[i]])
par(opar)

## We call a new graphic window
x11()
## A K-select analysis
acp <- dudi.pca(tab, scannf = FALSE, nf = 2)
kn <- kselect(acp, x$factor, x$weight,
 scannf = FALSE, nf = 2)

# use of the generic function scatter
scatter(kn)

# Displays the first factorial plane
kplot(kn)
kplot(kn, cellipse = 0, cpoint = 0)
kplot(kn, addval = FALSE, cstar = 0)

# this factorial plane can be compared with
# the other graph to see the rotation proposed by
# the analysis
graphics.off()

# Displays the first factorial axis
hist(kn)

# Displays the second factorial axis
hist(kn, xax = 2)

# Summary of the analysis
plot(kn)
## End(Not run)
```

`labcon` *Labelling Connected Features*

Description

This function attributes unique labels to pixels belonging to connected features on a map of class `asc`.

Usage

```
labcon(x)
```

Arguments

`x`	an object of class `asc`

Value

Returns a matrix of class `asc`, of type `"factor"`, with a number of levels equals to the number of connected components

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`as.asc`

Examples

```
data(puechabon)
hr <- getsahrlocs(puechabon$sahr, "hr")
u <- getkasc(hr, "Jean")
image(u)

## numbering of the connected components
p <- labcon(u)
nlevels(p)
image(p)

##  stores the first component
c1 <- p
c1[c1 != 1] <- NA
image(c1)

##  stores the second component
c2 <- p
c2[c2 != 2] <- NA
image(c2)
```

`lowres` *Reducing the Resolution of a Map*

Description

`lowres` is a generic function, having methods for the classes `asc` and `kasc`. It is used to reduce the resolution of the maps.

Usage

```
lowres(x, np = 2, ...)
```

Arguments

`x`	an object of class `asc` or `kasc`
`np`	a number giving the number of pixels to merge together (see below)
`...`	further arguments passed to or from other methods

Details

The function merges together squares of `np * np` pixels. For maps of type `"numeric"` (see `help(as.asc)`), the function averages the value of the variable. For maps of type `"factor"`, the function gives the most frequent level in the square of `np * np` pixels. When several levels are equally represented in the square of `np * np` pixels, the function randomly samples one of these levels.

Value

Returns an object of class `asc` or `kasc`.

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`as.asc`, `as.kasc`

Examples

```
data(puechabon)
kasc <- puechabon$kasc

## The initial image
image(kasc)

## The transformed image
m <- lowres(kasc, np = 4)
image(m)
```

`lynxjura` *Monitoring of Lynx*

Description

This data set stores the results of the monitoring of lynx in the French Jura between 1980 and 1999. These data have been collected by the Lynx Network of the french wildlife management office (Office national de la chasse et de la faune sauvage).

Usage

```
data(lynxjura)
```

Format

The list `lynxjura` has two components:

`map` an object of class `kasc` (see `help(as.kasc)`) that describes several variables on the study area: `forets` is the density of forests, `hydro` is the density of rivers, `routes` is the density of roads and `artif` is the distance from urbanized areas.

`locs` a data frame containing the locations of presence indices of the lynx. `X` and `Y` are the x and y coordinates, `Date` is the date of the collection of the indice and `Type` represents the type of data (`C`: alive lynx captured, `D`: attacks on livestock, `E`: prints or tracks, `F`: feces, `J`: hairs, `L`: corpse of lynx, `O`: sightings and `P`: attacks on wild prey).

Source

Vandel, J.M. (2001) *Répartition du Lynx (Lynx lynx) en France (Massif Alpin, Jurassien et Vosgien). Méthodologie d'étude et statut actuel.* Ecole Pratique des Haute Etudes de Montpellier II: Dissertation.

`mahasuhab` *Habitat Suitability Mapping with Mahalanobis Distances.*

Description

This function computes the habitat suitability map of an area for a species, given a set of locations of the species occurences (Clark et al. 1993). This function is to be used in habitat selection studies, when animals are not identified.

Usage

```
mahasuhab(kasc, pts, type = c("distance", "probability"))
```

Arguments

`kasc`	a raster map of class `kasc`
`pts`	a data frame with two columns, giving the coordinates of the species locations
`type`	a character string. Whether the raw `"distance"` should be returned, or rather the `"probability"` (see details).

Details

Let assume that a set of locations of the species on an area is available (gathered on transects, or during the monitoring of the population, etc.). If we assume that the probability of detecting an individual is independent from the habitat variables, then we can consider that the habitat found at these sites reflects the habitat use by the animals.

The Mahalanobis distance method has become more and more popular during the past few years to derive habitat suitability maps. The niche of a species is defined as the probability density function of presence of a species in the multidimensionnal space defined by the habitat variables. If this function can be assumed to be multivariate normal, then the mean vector of this distribution corresponds to the optimum for the species.

The function `mahasuhab` first computes this mean vector as well as the variance-covariance matrix of the niche density function, based on the value of habitat variables in the sample of locations. Then, the Mahalanobis distance from this optimum is computed for each pixel of the map. Thus, the smaller this distance is for a given pixel, and the better is the habitat in this pixel.

Assuming multivariate normality, Mahalanobis distances are approximately distributed as Chi-square with n-1 degrees of freedom, where n equals the number of habitat characteristics. If the argument `type = "probability"`, maps of these p-values are returned by the function. As such these are the probabilities of a larger Mahalanobis distance than that observed when x is sampled from a population whose mean is ideal.

Value

Returns a raster map of class `asc`.

Note

The computation of the Mahalanobis distances inverts the variance-covariance matrix of the niche density function (see `?mahalanobis`). It is therefore important that the habitat variables considered are not too correlated among each other. When the habitat variables are too correlated, the variance-covariance matrix is singular and cannot be inverted.

Note also that it is recommended to scale the variables before the computation, so that they all have the same variance, and therefore the same weight in the analysis (see examples below).

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

References

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

See Also

`as.asc` for further information on objects of class `asc`, `as.kasc` for additional information on objects of class `kasc`, `domain` for another method of habitat suitability mapping, and `mahalanobis` for information on the computation of mahalanobis distances.

Examples

```
## loads the data
data(lynxjura)
ka <- lynxjura$map
lo <- lynxjura$locs[,1:2]

## We first scale the maps
df <- kasc2df(ka)
pc <- dudi.pca(df$tab, scannf=FALSE)
tab <- pc$tab
ka <- df2kasc(tab, df$index, ka)

## habitat suitability mapping
hsm <- mahasuhab(ka, lo, type = "probability")
plot(hsm, main = "Habitat suitability map for the Lynx",
     plot.axes = { points(lo, pch = 16, cex=0.5)})
```

`managNAkasc` *"Cleaning" Objects of Class 'kasc'*

Description

An object of class `kasc` stores several maps in a data frame (one column per variable, and one row per pixel of the raster map). However, the features mapped are rarely rectangle-shaped, whereas the map are inevitably rectangles. Therefore, a lot of pixels of the maps do not contain data. The pixels of the map that do not contain data are `NA` in this data frame.
It often occurs that several variables are not mapped for exactly the same area (that is, some pixels are `NA` for some variables, and not for others). `managNAkasc` will set to `NA` all pixels that are not mapped for all variables.

Usage

```
managNAkasc(x)
```

Arguments

`x`	an object of class `kasc`

Value

Returns an object of class `kasc`.

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`as.kasc` for additional information on objects of class `kasc`.

`mcp` *Estimation of the Home Range Using the Minimum Convex Polygon Estimator*

Description

`mcp` computes the home range of several animals using the minimum convex polygon estimator.
`mcp.area` is used for home-range size estimation. This function requires the package `gpclib`.
`plot.hrsize` is used to display the home-range size estimated at various levels.

Usage

```
mcp(xy, id, percent = 95)
mcp.area(xy, id, percent = seq(20,95, by=5),
         unin = c("m", "km"),
         unout = c("ha", "km2", "m2"))
plot.hrsize(x, ...)
```

Arguments

`xy`	a data frame containing the coordinates of the relocation of the monitored animals
`id`	a factor giving the identity of the animal for each relocation
`percent`	100 minus the proportion of outliers to be excluded from the computation
`unin`	the units of the relocations coordinates. Either `"m"` (default) for meters or `"km"` for kilometers
`unout`	the units of the output areas. Either `"m2"` for square meters, `"km2"` for square kilometers or `"ha"` for hectares (default)
`x`	an objet of class `hrsize` returned by the function `mcp.area`, or `kernel.area` (see the help page of this function)
`...`	additionnal arguments to be passed to the function `plot`

Details

This function computes the Minimum Convex Polygon estimation after the removal of `(100 minus percent)` percent of the relocations the farthest away from the barycenter of the home range (computed by the arithmetic mean of the coordinates of the relocations for each animal).

Value

`mcp` returns an object of class `area`, with one polygon per level of the factor ID.
`mcp.area` returns a data frame of class `hrsize`, with one column per animal and one row per level of estimation of the home range

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

References

Mohr, C.O. (1947) Table of equivalent populations of north american small mammals. *The American Midland Naturalist*, **37**, 223-249.

See Also

`chull`, `plot.area` to have a graphical display of the home ranges, `as.area` for additionnal information on the class `area`, and `area2dxf` for further exportation toward a GIS. `s.chull` for another way to display MCP

Examples

```
data(puechabon)
locs <-  puechabon$locs

cp <- mcp(locs[,4:5], locs[,1])

## Plot the home ranges
opar <- par(mar = c(0,0,0,0))
area.plot(cp)

## ... And the relocations
points(locs[,4:5], pch = 16, col = as.numeric(locs[,1]))
par(opar)

## Computation of the home-range size:
if (require(gpclib)) {
  (cuicui1 <- mcp.area(locs[,4:5], locs[,1]))
  plot(cuicui1)
}
```

`mcp.rast` *Converts a Polygon to Raster*

Description

`mcp.rast` converts a polygon map to a raster map of class `asc`.

Usage

```
mcp.rast(poly, w)
```

Arguments

`poly`	a data frame with 2 columns giving the coordinates of a polygon object
`w`	an object of class `kasc`, or of class `asc`

Value

Returns an object of class `asc`.

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`hr.rast`

Examples

```
data(puechabon)
toto <- puechabon$kasc
loc <- puechabon$locs

## gets the coordinates of the relocations for the wild boar #1
wb1 <- loc[loc$Name == "Chou",]
wb1 <- cbind(wb1$X, wb1$Y)
nbpol <- chull(wb1)
xycoord <- wb1[nbpol,]

## rasterization of wb1
tutu <- mcp.rast(xycoord, toto)
image(tutu)

polygon(xycoord, lwd = 2)
```

`morphology` *Morphology: Erosion or Dilatation of Features on a Raster Map*

Description

`morphology` performs morphological operations on images of class `asc`.

Usage

```
morphology(x, operation = c("erode", "dilate"), nt = 5)
```

Arguments

`x`	a matrix of class `asc`
`operation`	a character string indicating the operation to be processed: either `"erode"` or `"dilate"`
`nt`	the number of times that the operation should be processed

Value

Returns a matrix of class `asc`, containing 1 when the pixel belong to one feature of the image and `NA` otherwise (see examples).

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`as.asc` for further information on objects of class `asc`

Examples

```
data(puechabon)
a <- getkasc(puechabon$kasc,"Elevation")

## dilatation
toto1 <- morphology(a, operation = "dilate", nt = 1)
toto2 <- morphology(a, operation = "dilate", nt = 2)
toto3 <- morphology(a, operation = "dilate", nt = 3)
toto5 <- morphology(a, operation = "dilate", nt = 5)
image(toto5, col = "red")
image(toto3, add = TRUE, col = "orange")
image(toto2, add = TRUE, col = "yellow")
image(toto1, add = TRUE, col = "green")
image(a, add = TRUE)

## erosion
image(a, col = 1)
for (i in 1:19) {
  toto <- morphology(a, operation = "erode", nt = i)
  image(toto, add = TRUE, col = palette()[(i+1)%%8])
}
```

`niche.test` *Monte-Carlo Test on Parameters of the Ecological Niche*

Description

`niche.test` tests for the significance of two parameters of the ecological niche of a species (marginality and tolerance), using Monte-Carlo methods. This is a bivariate test.

Usage

```
niche.test(kasc, points, nrep = 999, o.include = TRUE, ...)
```

Arguments

`kasc`	a raster map of class `kasc`
`points`	a data frame with two columns, giving the coordinates of the species locations
`nrep`	the number of permutations
`o.include`	logical, passed to `biv.test`. If `TRUE`, the origin is included in the plot
`...`	further arguments passed to `biv.test`

Details

`niche.test` tests the significance of two parameters describing the ecological niche: the marginality (squared length of the vector linking the average available habitat conditions to the average used habitat conditions in the ecological space defined by the habitat variables), and the tolerance (inertia of the niche in the ecological space, i.e. the sum over all variables of the variance of used pixels).

At each step of the randomisation procedure, the test randomly allocates the n points in the pixels of the map. The marginality and the tolerance are then recomputed on this randomised data set.

Actual values are compared to random values with the help of the function `biv.test`.

Value

Returns a list containing the following components:

`dfxy`	a data frame with the randomized values of marginality (first column) and tolerance (second column).
`obs`	the actual value of marginality and tolerance.

Warning

ă `biv.test` uses the function `kde2d` of the package `MASS`.

Author(s)

Mathieu Basille ⟨basille@biomserv.univ-lyon1.fr⟩
Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`biv.test` for more details on bivariate tests. `histniche` for the histograms of the variables of the niche.

Examples

```
## Not run:
data(lynxjura)

## We keep only "wild" indices.
tmp=lynxjura$loc[,4]!="D"
niche=niche.test(lynxjura$map,
                 lynxjura$locs[tmp, c("X", "Y")],
                 side = "bottom")
names(niche)
## End(Not run)
```

`perarea` *Compute Areas and Perimeters of Objects of Class "area"*

Description

`perarea` computes the perimeters of polygons in objects of class `area`.
`ararea` computes the areas of polygons in objects of class `area`.

Usage

```
perarea(x)
ararea(x)
```

Arguments

`x`	object of class `area`

Value

a vector.

Warning

`ararea` requires the package `gpclib`.

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`as.area`, `plot.area`

Examples

```
data(puechabon)
locs <-  puechabon$locs

cp <- mcp(locs[,4:5], locs[,1])

perarea(cp)

if (require(gpclib))
   ararea(cp)
```

`pheasant` *Radio-Tracking of Pheasants*

Description

This data set describes the use and availability of 5 habitat types for 13 pheasants monitored using radio-tracking.

Usage

```
data(pheasant)
```

Format

This list has three components:

- **`studyarea`** a data frame giving the proportion of each habitat type (columns) on the study area. These habitat types are `Scrub`, `Broadleaf`, `Coniferous`, `Grassland` and `Crop`. These proportions are repeated by rows, for all animals (rows)
- **`mcp`** a data frame giving the proportion of each habitat type (columns) in the home range of each animal (rows)
- **`locs`** a data frame giving the proportion of relocations of each animal (rows) reported in 3 of the 5 habitat types (columns). Indeed, `Coniferous` and `Crops` were not used by most of the animals.

Source

Aebischer, N. J., Robertson, P. A. and Kenward, R. E. (1993) Compositional analysis of habitat use from animal radiotracking data. *Ecology*, **74**, 1313–1325.

`plot.area` *Graphical Display of Objects of Class "area"*

Description

`plot.area` allows a graphical display of objects of class `"area"`.

Usage

```
plot.area(x, which = levels(x[,1]),
          colpol = rep("green", nlevels(x[, 1])),
          colborder = rep("black", nlevels(x[, 1])),
          lwd = 2, add = FALSE, ...)
```

Arguments

`x`	an object of class `"area"`
`which`	a vector of character indicating the polygons to be plotted
`colpol`	a vector of the color for filling the polygon. The default, `NA`, is to leave polygons unfilled
`colborder`	a vector of the color to draw the border. The default. Use `border = NA` to omit borders
`lwd`	the border width, a **positive** number
`add`	logical. if `TRUE`, the polygons are added to a previous plot
`...`	additional arguments to be passed to the generic function `plot`

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`area.plot` for another way to display objects of class `"area"`

Examples

```
## Loading the relocations of wild boars
## monitored using radio-tracking
data(puechabon)
locs <-  puechabon$locs
el <- getkasc(puechabon$kasc, "Elevation")

## Estimation of the MCP home ranges of the animals
cp <- mcp(locs[,4:5], locs[,1])

## Use of plot.area to display the results
plot(cp)

## different colors:
color <- c("red", "blue", "green", "yellow")
lev <- levels(cp[,1])
image(el)
plot(cp, colpol = color, add = TRUE)
legend(697198, 3165529, legend = lev, fill = color)

## or:
image(el)
plot(cp, colborder = color, colpol = NA, add = TRUE)
legend(697198, 3165529, legend = lev, fill = color)

## plots one animal
image(el)
plot(cp, which = "Brock", add = TRUE)
```

`plot.sahrlocs` *Exploratory Analysis of Habitat Selection*

Description

`plot.sahrlocs` applies the function `widesII` or `widesIII` for each variable in the object of class `sahrlocs`, and the results are stored in a list. Then, the function `plot.wi` is applied to each component of the list. This allows to investigate habitat selection by animals at several scales for design II and III data.

Usage

```
plot.sahrlocs(x, ani = names(x$hr), var = names(x$sa),
              type = c("hr.in.sa", "locs.in.hr", "locs.in.sa"),
              ncla = 4, ylog = FALSE, caxis = 0.7, clab = 0.7,
              errbar = c("SE", "CI"), alpha = 0.05, draw = TRUE, ...)
```

Arguments

`x`	an object of class `sahrlocs`
`ani`	a character vector. This vector contains the names of the animals in `x` for which habitat selection should be displayed. At least two animals are required
`var`	a character vector. This vector contains the names of the variables in `x` for which habitat selection should be displayed
`type`	a character string. Type of habitat selection that should be investigated. If `"hr.in.sa"`, the selection of the home-range within the study area is displayed. If `"locs.in.sa"`, the selection of the relocations within the study area is displayed. If `"locs.in.hr"`, the selection of the relocations within the home range is displayed
`ncla`	numeric variables are converted to factors. This parameter controls the number of classes of these factors
`ylog`	logical. If `TRUE`, the selection ratios are plotted on a log scale
`caxis`	character size on axes to be passed to `par("cex.axis")`
`clab`	character size of axes labels to be passed to `par("cex.lab")`
`errbar`	a character string. Type of error bars: either `"CI"` for confidence intervals or `"SE"` for standard errors
`alpha`	the alpha-level for the tests
`draw`	logical. If `FALSE`, no plot is drawn
`...`	further arguments to be passed to the function `plot.wi`

Value

The function returns a list of objects of class `wiII` or `wiIII` (one component per animal).

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

References

Manly B.F.J., McDonald L.L., Thomas, D.L., McDonald, T.L. & Erickson, W.P. (2003) *Resource selection by animals - Statistical design and Analysis for field studies. Second edition.* London: Kluwer academic publishers.

See Also

`widesII` and `widesIII` for further information on objects of class `wiII` and `wiIII`, `as.sahrlocs` for further information on objects of class `sahrlocs`.

Examples

```
data(puechabon)
sahr <- puechabon$sahr
toto <- plot(sahr)




### Note that the wild boars named Brock and Calou
### have only one herbaceous cover class available
### (the second one), and they use it exclusively.
### So they have identical
### selection ratios (that's why the curve of Brock
### does not appear: it is hidden behind the curve
### of the boar named Calou).

toto
toto$Elevation
```

`predict.enfa` *Habitat Suitability Maps Built from the ENFA*

Description

`predict.enfa` computes habitat suitability maps using the Ecological-Niche Factor Analysis and the Mahalanobis distances method.

Usage

```
predict.enfa(object, index, attr, nf, ...)
```

Arguments

`object`	an object of class `enfa`
`index`	an integer vector giving the position of the rows of `tab` in the initial object of class `kasc`.
`attr`	an object of class `kasc` or `mapattr`.
`nf`	the number of axes of specialization kept for the predictions. By default, all axes kept in `object` are used
`...`	further arguments passed to or from other methods

Details

The predictions are based on the position of the niche defined by the ENFA within the multidimensional space of environmental variables. The ENFA produces row coordinates for each pixel, which are used with the function `mahalanobis`. For each pixel, this function computes the Mahalanobis distances from the barycentre of the niche.

Actually, the function `predict.enfa` is identical to the function `mahasuhab`, except that the habitat suitability map is computed using the axes of the ENFA, instead of the raw data.

Value

Returns a raster map of class `kasc`.

Author(s)

Mathieu Basille ⟨basille@biomserv.univ-lyon1.fr⟩

References

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

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

See Also

`mahalanobis` for information on the computation of Mahalanobis distances. `mahasuhab` for more details on the computation of habitat suitability maps using the Mahalanobis distances.

Examples

```
## Not run:
data(lynxjura)

map <- lynxjura$map

## We keep only "wild" indices.
tmp <- lynxjura$loc[,4] != "D"
locs <- lynxjura$locs[tmp, c("X","Y")]
dataenfa1 <- data2enfa(map, locs[tmp, c("X","Y")])

(enfa1 <- enfa(dataenfa1$tab, dataenfa1$pr,
               scannf = FALSE))

## Compute the prediction
pred <- predict(enfa1, dataenfa1$index, dataenfa1$attr)
image(pred)
contour(pred, col="green", add=T)
points(locs, col = "red", pch = 16)
## Lighter areas are the most preferred areas
## End(Not run)
```

`puechabon` *Radio-Tracking Data of Wild Boar*

Description

This data set stores the results of the monitoring of 4 wild boars at Puéchabon (Mediterranean habitat, South of France). These data have been collected by Daniel Maillard (Office national de la chasse et de la faune sauvage).

Usage

```
data(puechabon)
```

Details

The list `puechabon` has three components:
`puechabon$kasc` is an object of class `kasc` (see `help(as.kasc)`) that describes several variables on the study area.
`locs` is a data frame of the relocations of the wild boar resting sites in summer. Information on wild boars is provided by factors `Name`, `Sex`, `Age`.
`sahr` is the associated object of class `sahrlocs`. The home ranges were estimated with buffers including all pixels within 500 m of a boar relocation.

References

Maillard, D. (1996). *Occupation et utilisation de la garrigue et du vignoble méditerranéens par le Sanglier*. Université d'Aix-Marseille III: PhD thesis.

`puechcirc` *Movements of wild boars tracked at Puechabon*

Description

This data set is an object of class `traj`, giving the results of the monitoring of 2 wild boars by radio-tracking at Puéchabon (Mediterranean habitat, South of France). These data have been collected by Daniel Maillard (Office national de la chasse et de la faune sauvage), and correspond to the activity period of the wild boar (during the night, when the animals forage. The data set `puechabon` describes the resting sites).

Usage

```
data(puechcirc)
```

Format

This data frame has 204 observations, and 5 variables:

id a factor with levels `CH93` and `JE93`, giving the name of the animals

x a vector giving the x coordinates of the relocations

y a vector giving the y coordinates of the relocations

date a POSIXct giving the date of collection of the relocations

burst a factor with levels `CH930803`, `CH930824`, `CH930827`, `JE930819`, and `JE930827`, giving the ID of the circuits.

Source

Maillard, D. (1996). *Occupation et utilisation de la garrigue et du vignoble méditerranéens par le Sanglier*. Université d'Aix-Marseille III: PhD thesis.

`puechdesIII` *Habitat Selection by the Wild Boar at Puechabon*

Description

This data set contains two data frames describing the use and the availability of 3 elevation classes for 6 wild boars (*Sus scrofa* L.) monitored using radio-tracking at Puéchabon (South of France). These data have been collected by Daniel Maillard (Office national de la chasse et de la faune sauvage).

Usage

```
data(puechdesIII)
```

Details

The list `puechdesIII` has two components:
The data frame `used` describes the number of telemetry relocations for each of the 6 animals in each of the three elevation classes.
The data frame `available` describes a sample of random points placed in the areas available to these wild boars (a buffer area of 200 m around the relocations).

Source

Maillard, D. (1996) *Occupation et utilisation de la garrigue et du vignoble méditerranéens par le Sanglier*. Université d'Aix-Marseille III: PhD thesis.

`rand.kselect` *Test of the Third-Order Habitat Selection*

Description

`rand.kselect` tests whether the marginality vector of animals is significantly larger than what is expected under the hypothesis of random habitat use (third-order habitat selection: selection by the animals of the relocations within their home range; the habitat availability is measured for each animal). The effect of each variable on individual marginality is also tested. Finally, the pertinence of a K-select analysis is also tested. This is a randomisation test. The alpha-level of the tests is ajusted using the Bonferroni inequality.

Usage

```
rand.kselect(dudi, factor, weight, nrep = 200, alpha = 0.05, ewa = FALSE)
print.rand.kselect(x, ...)
```

Arguments

`dudi`	an object of class `dudi`.
`factor`	a factor defining the animals identity
`weight`	a weight vector of integer values (number of relocations counted in each resource unit in row of the object dudi)
`nrep`	the number of repetitions of the test
`alpha`	the alpha level for the tests.
`ewa`	logical. If `TRUE`, uniform weights are given to all animals in the analysis. If `FALSE`, animal weights are given by the proportion of relocations of each animal (i.e. an animal with 10 relocations has a weight 10 times lower than an animal with 100 relocations)
`x`	an object of the class `rand.kselect`.
`...`	further arguments to be passed to the generic function `print`

Details

This test is carried out by simulating a random use of space by animals. `rand.kselect` is closely related to the function `kselect` (same arguments).

At each step of the randomisation procedure, and for each animal, the test randomly allocates the nk relocations (where nk is the sum of the weight vector for the animal k) in the Ik pixels available to this animal (where Ik is the length of the weight vector for animal k).

The length of the marginality vector is recomputed at each step of the randomisation procedure and for each animal. The effect of each variable on the use of pixels by each animal is measured by the criterion "(average habitat variable j used by animal i) minus (average habitat variable j available to animal i)". Finally the value of the first eigenvalue of the K-select analysis provides a criterion to test the pertinence of the K-select analysis.

All these values are then compared to the observed values to assess the significance of theses effects.

Value

Returns an object of class `rand.kselect`. This list has three components:

`global`	a vector of length 2 giving the results of the randomisation procedure for the first eigenvalue of the K-select analysis.
`marg`	a matrix giving the significance of the marginality of each animal.
`per.ind`	a list giving the results of the randomisation test for the coordinates of the marginality vector for each animal on each habitat variable.
`alpha`	the alpha level of the tests.

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

References

Calenge, C., Dufour, A.B. and Maillard, D. (*submitted*). K-select analysis, a new method to analyse habitat selection in radio-tracking studies.

See Also

`kselect` to perform a K-select analysis.

Examples

```
## Not run:
## Loads the data
data(puechabon)
sahr <- puechabon$sahr

## prepares the data for the kselect analysis: only two variables are kept
x <- sahrlocs2kselect(sahr)
tab <- x$tab
dud <- dudi.mix(tab, scannf = FALSE, nf = 2)

## the randomisation tests
## be patient, this can be very long on some machines
(te <- rand.kselect(dud, x$factor, x$weight, nrep = 50))

## End(Not run)
```

`randtest.enfa` *Randomisation Test for the Ecological Niche Factor Analysis*

Description

`randtest.enfa` performs a randomisation test for the Ecological Niche Factor analysis (ENFA).

Usage

```
randtest.enfa(xtest, nrepet = 999, ...)
```

Arguments

`xtest`	an object of class `enfa`
`nrepet`	the number of iterations for the randomisation test
`...`	further arguments to be passed to the generic function `randtest`

Details

This test is carried out by simulating a random distribution of the species occurrences in the pixels of a map.

At each step of the randomisation procedure, the test randomly allocates the nk occurrences (where nk is the sum of the occurrence vector `pr` of the object of class `enfa`) in the Ik pixels of the focus area (where Ik is the length of this occurrence vector).

At each step of the procedure, the first eigenvalue of the ENFA performed on the randomised data set is recomputed. This value provides a criterion to test the pertinence of the ENFA analysis.

Value

returns a list of class `randtest`

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

References

Manly, B.F.J. (1997) *Randomization, Bootstrap and Monte Carlo Methods in Biology*. London: Chapman & Hall.

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

See Also

`enfa`

Examples

```
## Not run:
data(chamois)
locs <- chamois$locs
map <- chamois$map
map$Vegetation <- NULL
en <- enfa(map, locs, scannf=F)
(tutu <- randtest(en))
plot(tutu)
## End(Not run)
```

`rotxy` *Internal Functions for Package 'adehabitat'*

Description

Internal functions for the package `adehabitat`.

Usage

Details

These are not to be called by the user.

`sahrlocs2kselect` *Preparation of K-Select Analysis*

Description

`sahrlocs2kselect` converts an object of class sahrlocs into a list that contains the arguments needed for a kselect analysis (see `help(kselect)`).

Usage

```
sahrlocs2kselect(sahr)
```

Arguments

`sahr`	an object of class `sahrlocs`

Value

Returns a list with the following components:

`tab`	a data frame giving the habitat composition of the home range of animals
`factor`	a factor giving for each row of `tab` the name of the corresponding animal
`weight`	a numeric vector giving for each row of `tab` the number of relocations numbered in this cell of the raster map

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

References

Calenge, C., Dufour, A.B. and Maillard, D. (*submitted*). K-select analysis, a new method to analyse habitat selection in radio-tracking studies.

See Also

`as.sahrlocs` for additional information on the objects of class `sahrlocs`, `kselect` for a K-select analysis, and `kasc2df` for additional information on the index component of the output.

Examples

```
data(puechabon)
sahr <- puechabon$sahr
s <- sahrlocs2kselect(sahr)

# 1. PCA of the dataset
pc <- dudi.mix(s$tab, scannf = FALSE)

# 2. k-select analysis
kn <- kselect(pc, s$factor, s$weight, scannf = FALSE)

scatter(kn)
```

`sahrlocs2niche` *OMI Analysis of Radio-Tracking Data*

Description

`sahrlocs2niche` converts an object of class `sahrlocs` into a list that contains the arguments needed for a niche analysis (function `niche` in package `ade4`).

Usage

```
sahrlocs2niche(x, ani = names(x$hr), var = names(x$sa), used = c("hr",
"locs"))
```

Arguments

`x`	an object of class `sahrlocs`.
`ani`	a character vector giving the name of the animals in the analyses
`var`	a character vector giving the name of the variables in the analyses
`used`	a character string. If `"hr"`, the cells of the raster map that are considered used by the animals are taken from the `hr` component of the object `sahrlocs`, if `"locs"`, the cells of the raster map that are considered used by the animals are taken from the `locs` component (see `as.sahrlocs`))

Value

Returns a list with three components:

`index`	a vector of integer giving the index of the rows of `x$sa` kept for the analysis (this component may then be used with the function `df2kasc`).
`tab`	the table on which the `dudi` analysis will be processed.
`y`	a table giving the weights of the niche analysis.

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

References

Dolédec, S., Chessel, D. and Gimaret, C. (2000) Niche separation in community analysis: a new method. *Ecology*, **81**, 2914–1927.

See Also

`as.sahrlocs` for additional information on the objects of class `sahrlocs`, `niche` for an OMI analysis, and `kasc2df` for additional information on the `index` component of the output.

Examples

```
data(puechabon)
sahr <- puechabon$sahr

s <- sahrlocs2niche(sahr, used = "locs")

# 1. dudi.mix of the dataset
pc <- dudi.mix(s$tab, scannf = FALSE)

# 2. niche analysis
n <- niche(pc, s$y, scannf = FALSE)

plot(n)
```

`scatter.enfa` *Scatter Plot of the Results of the ENFA*

Description

Performs the scatter diagrams of objects of class `enfa`.

Usage

```
scatter.enfa(x, xax = 1, yax = 2, pts = FALSE, nc = TRUE,
             percent = 95, clabel = 1, side = c("top", "bottom", "none"),
             Adensity, Udensity, Aangle, Uangle, Aborder, Uborder,
             Acol, Ucol, Alty, Ulty, Abg, Ubg, Ainch, Uinch, ...)
```

Arguments

`x`	an object of class `enfa`
`xax`	the column number for the x-axis
`yax`	the column number for the y-axis
`pts`	logical. Whether the points should be drawn. If `FALSE`, minimum convex polygons are displayed
`nc`	whether or not the niche center should be displayed

`percent`	100 minus the proportion of outliers to be excluded from the computation of the minimum convex polygons
`clabel`	a character size for the columns
`side`	if `"top"`, the legend of the kept axis is upside, if `"bottom"` it is downside, if `"none"` no legend
`Adensity`	the density of shading lines, in lines per inch, for the available pixels polygon. See `polygon` for more details
`Udensity`	the density of shading lines, in lines per inch, for the used pixels polygon. See `polygon` for more details
`Aangle`	the slope of shading lines, given as an angle in degrees (counter-clockwise), for the available pixels polygon
`Uangle`	the slope of shading lines, given as an angle in degrees (counter-clockwise), for the used pixels polygon
`Aborder`	the color to draw the border of the available pixels polygon. See `polygon` for more details
`Uborder`	the color to draw the border of the used pixels polygon. See `polygon` for more details
`Acol`	the color for filling the available pixels polygon. if `pts==FALSE`, the color for the points corresponding to available pixels
`Ucol`	the color for filling the used pixels polygon. if `pts==FALSE`, the color for the points corresponding to used pixels
`Alty`	the line type for the available pixels polygon, as in `par`.
`Ulty`	the line type for the used pixels polygon, as in `par`.
`Abg`	if `pts==TRUE`, background color for open plot symbols of available pixels
`Ubg`	if `pts==TRUE`, background color for open plot symbols of used pixels
`Ainch`	if `pts==TRUE`, heigth in inches of the available pixels
`Uinch`	if `pts==TRUE`, heigth in inches of the largest used pixels
`...`	further arguments passed to or from other methods

Details

`scatter.enfa` displays a factorial map of pixels, as well as the projection of the vectors of the canonical basis multiplied by a constant of rescaling. The kept axes for the plot are specified in a corner.

Author(s)

Mathieu Basille ⟨basille@biomserv.univ-lyon1.fr⟩

See Also

`enfa`, `scatter`

Examples

```
## Not run:
## Not run:
data(lynxjura)

map <- lynxjura$map

## We keep only "wild" indices.
tmp <- lynxjura$locs[,4]!="D"
locs <- lynxjura$locs[tmp, c("X","Y")]

## We perform a square root transformation
## of the variable to normalize it
map[,4] <- sqrt(map[,4])

## We perform the ENFA
(enfa1 <- enfa(map, locs[tmp, c("X","Y")],
                scannf = FALSE))
scatter(enfa1)
## End(Not run)
```

`schoener` *Compute Schoener's ratio*

Description

`schoener` computes the Schoener's ratio on radio-tracking data.
`schoener.rtest` performs a randomization test of the equality of the Schoener's ratio to 2

Usage

```
schoener(tr, keep, byburst = TRUE)
schoener.rtest(tr, keep, byburst = TRUE, nrep = 500)
```

Arguments

`tr`	an object of class `traj`
`keep`	a vector of length 2, giving the lower and the upper bound of the time interval for which a pair of relocations is considered in the computation of t$\hat{2}$ (see details). These values are given in seconds.
`byburst`	logical. If `TRUE`, the Schoener's ratio is computed by burst. If `FALSE`, the ratio is computed by animal.
`nrep`	the number of randomisations of the test.

Details

The Schoener's ratio is a measure of time-autocorrelation in the data. This ratio is computed as the squared mean distance between "neighbour" relocations (t$\hat{2}$) divided by the squared mean distance between the relocations and their barycenter (r$\hat{2}$). The theoretical value of this ratio under the hypothesis of independance of the relocations is 2.

Swihart and Slade (1985) consider as neighbour two successive relocations. However, the Schoener's ratio computed in this way makes sense biologically only if the relocations are equally spaced in time. However, as indicated by these authors, "such a data set probably is the exception rather than the rule because many problems may arise in taking a locational reading at a specified time".

In this function, we define as "neighbour" two relocations (not necessarily successive relocations) separated by a time interval comprised within the bounds specified in the vector `keep` (in seconds). For example, if `keep = c(60, 300)`, all relocations separated by a time interval comprised between 1 and 5 minutes are considered in the computation. Thus, the total number of pairs of relocations m taken into account in the computation may be larger than the number of relocations n (m can be at most equal to n*(n-1)/2).

Value

returns an object of class `schoener`.

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

References

Swihart, R.K. and Slade, N.A. (1985) Testing for independence of observations in animal movements. *Ecology*, **66**, 1176–1184.

Schoener, T.W. (1981) An empirically based estimate of home range. *Theoretical Population Biology*, **20**, 281–325.

Examples

```
data(puechcirc)
puechcirc$date[1:10]

## Relocations are taken every 10 minutes
## For example we consider relocations as
## neighbour when they are separated by a time
## interval comprised between 5 and 15 minutes
schoener(puechcirc, keep = c(5*60, 15*60))
## Not run:
schoener.rtest(puechcirc, keep = c(5*60, 15*60))
## End(Not run)
```

`setmask` *Applies a Mask on Objects of Class 'asc' or 'kasc'*

Description

Applies a mask on objects of class `asc` or `kasc`. In other words, the function creates an object of class `asc` or `kasc`, with `NA` for all pixels `NA` on the masking map.

Usage

```
setmask(x, mask)
```

Arguments

`x`	an object of class `asc` or `kasc`
`mask`	an object of class `asc`

Value

Returns an object of class `asc` or `kasc`

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`import.asc` for additionnal information on objects of class `asc`.

Examples

```
data(puechabon)
kasc <- puechabon$kasc
image(kasc)
elev <- getkasc(kasc, "Elevation")
slope <- getkasc(kasc, "Slope")

## ma is the mask: only areas with elevation > 250 m
## are kept
ma <- elev
ma[ma < 250] <- NA
ma <- getascattr(elev, ma)
image(ma)

## The mask is applied on maps of slope
slp <- setmask(slope, ma)
image(slp)

## The mask is applied on all maps in kasc
im <- setmask(kasc, ma)
image(im)
```

`speed` *Computes the Speed Between Successive Relocations of an Animal*

Description

`speed` measures the speed between successive relocations of animals, using objects of class `traj`.

Usage

```
speed(x, id = levels(x$id), burst = levels(x$burst), date = NULL,
      units = c("seconds", "hours", "days"))
```

Arguments

`x`	an object of class `traj`
`id`	a character vector giving the identity of the animals for which the speed is to be computed
`burst`	a character vector giving the identity of the circuits for which the speed is to be computed (see `as.traj`)
`date`	a vector of class `POSIXct` of length 2 (beginning, end) delimiting the period of interest
`units`	a character string. It determines how the speeds are computed. For example, if the coordinates are given in meters, and if `units = "seconds"`, speeds are returned in meters per second.

Value

Returns a data frame with the following components:

`id`	the identity of the animal
`x`	if the speed is computed between the relocation 1 and 2, the x coordinate of the relocation 2.
`y`	if the speed is computed between the relocation 1 and 2, the y coordinate of the relocation 2.
`date`	a vector of class `POSIXct`, giving the date of relocation 2.
`burst`	the identity of the circuit
`sp.x`	the computed speed of the animal in the x direction
`sp.y`	the computed speed of the animal in the y direction
`speed`	the computed speed of the animal on the plane.
`dt`	the duration between the two relocations (in the units given by the parameter `units`).

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`as.traj`

Examples

```
#### Computes the speed for each wild boar
#### monitored at Puechabon
data(puechcirc)
puechcirc
plot(puechcirc)
```

```
sp <- speed(puechcirc)
sp[1:4,]
```

`squirrel` *Radio-Tracking Data of Squirrels*

Description

This data set describes the use and availability of 5 habitat types for 17 squirrels monitored using radio-tracking.

Usage

```
data(squirrel)
```

Format

This list has three components:

- **`studyarea`** a data frame giving the proportion of each habitat type (columns) on the study area. These proportions are repeated by rows, for all animals
- **`mcp`** a data frame giving the proportion of each habitat type (columns) in the home range of each animal (rows)
- **`locs`** a data frame giving the proportion of relocations of each animal (rows) reported in each habitat type (columns).

Source

Aebischer, N. J., Robertson, P. A. and Kenward, R. E. (1993) Compositional analysis of habitat use from animal radiotracking data. *Ecology*, **74**, 1313–1325.

`storemapattr` *Store attributes of maps of class asc and kasc*

Description

`storemapattr` stores attributes of maps of class `asc` and `kasc` in an object of class `mapattr`.

Usage

```
storemapattr(x)
```

Arguments

`x`	An object of class `asc` or `kasc`

Details

This function is essentially used by programmers in functions dealing with maps of class `asc` or `kasc`. The function `getXYcoords` is an example of function using `storemapattr`.

Value

Returns an object of class `mapattr`.

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`as.kasc` and `import.asc` for additional information on the classes `kasc` and `asc`.

Examples

```
data(puechabon)
(kasc <- puechabon$kasc)

(toto <- storemapattr(kasc))
```

`subsetmap` *Storing a Part of a Map*

Description

`subsetmap` is a generic function. It has methods for the classes `asc` and `kasc`. It is used to store a part of any given map into an other object.

Usage

```
subsetmap(x, xlim = NULL, ylim = NULL, ...)
```

Arguments

`x`	an object of class `asc` or `kasc`
`xlim`	numerical vector of length 2. The x limits of the rectangle including the new map
`ylim`	numerical vector of length 2. The y limits of the rectangle including the new map
`...`	further arguments passed to or from other methods

Details

If `xlim` or `ylim` are not provided, the function asks the user to click on the map to delimit the lower left corner and the higher right corner of the new map (see Examples).

Value

Returns an object of class `asc` or `kasc`

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

See Also

`as.asc`, `as.kasc`

Examples

```
data(puechabon)
kasc <- puechabon$kasc
el <- getkasc(kasc, "Elevation")

## limits of the new map:
xl <- c(701561, 704017)
yl <- c(3160560, 3162343)

## computation of the new map:
su <- subsetmap(el, xlim = xl, ylim = yl)

## Display
opar <- par(mar = c(0,0,0,0))
layout(matrix(c(1,1,1,1,1,1,1,1,2), byrow = TRUE, ncol = 3))
image(el, axes = FALSE)
polygon(c(xl[1], xl[2], xl[2], xl[1]),
        c(yl[1], yl[1], yl[2], yl[2]))
image(su, axes = FALSE)
box()

par(opar)
par(mfrow = c(1,1))

### Gets this part for the whole kasc object
m <- subsetmap(kasc, xlim = xl, ylim = yl)
image(m)

## Not run:
 ## Interactive example
 su <- subsetmap(kasc)

 image(su)
## End(Not run)
```

`vanoise` *Habitat Use by Three Species of Galliformes in the Vanoise National Parc*

Description

This data frame describes the habitat use by the Black Grouse (*Tetrao tetrix*), the Rock Partridge (*Alectoris graeca*) and the Rock Ptarmigan (*Lagopus mutus*), in the Vanoise National Park (French Alps), from 1990 to 2000.

Usage

```
data(vanoise)
```

Format

This data frame has 3110 rows and eight columns describing the habitat composition for each occurrence of three species of Galliformes. For each located occurrence (in rows), the employees of the national park have noted: the species, the elevation (in metres), the aspect (8 classes), the habitat type (7 categories, `FR` means "fallen rocks") and the date (season, day, month and year).

Source

Calenge, C., Martinot, J.P. and Lebreton, P. (2003) Ecological niche separation among mountain Galliformes in the Vanoise National Parc. *Game and Wildlife Science*, 20, 259-285.

`wi` *Computation of Selection Ratios for Habitat Selection Studies.*

Description

These functions compute the resource selection ratios (wi) for design I, II and III data types, with resources defined by several categories. Basic tests are also provided.

Usage

```
widesI(u, a, avknown = TRUE, alpha = 0.05)
widesII(u, a, avknown = TRUE, alpha = 0.05)
widesIII(u, a, avknown = TRUE, alpha = 0.05)
print.wiI(x, ...)
print.wiII(x, ...)
print.wiIII(x, ...)
plot.wi(x, caxis = 0.7, clab = 1, ylog = FALSE, errbar = c("CI", "SE"),
        main = "Manly selectivity measure", noorder = TRUE, ...)
```

Arguments

`u`	for `widesI`, a vector with named elements describing the sample of used resource units. For `widesII` and `widesIII` a matrix or a data frame giving the number of used resource units for each animal (in rows) in each resource category (in columns)
`a`	for `widesI` and `widesII`, a vector with named elements describing the sample or the proportion of available resource units. For `widesIII` a matrix or a data frame giving the number or the proportion of available resource units for each animal (in rows) in each resource category (in columns)
`avknown`	logical. `TRUE` if the available proportions are known, and `FALSE` if they are estimated
`alpha`	the threshold value for the tests and confidence intervals
`x`	an object of class `wi`
`caxis`	character size on axes to be passed to `par("cex.axis")`
`clab`	character size of axes labels to be passed to `par("cex.lab")`
`ylog`	logical. If `TRUE`, the selection ratios are plotted on a log scale
`errbar`	a character string. Type of error bars: either `"CI"` for confidence intervals or `"SE"` for standard errors

`main`	a character string. The title of the graph
`noorder`	logical. If `TRUE`, the habitat categories are ordered on the graph in decreasing order of their preference. If FALSE, they are not ordered (i.e. they are in the same order as the columns in `used` and `available`
`...`	additionnal arguments to be passed to the function `plot`

Details

`widesI` may be used to explore resource selection by animals, when designs I are involved (habitat use and availability are measured at the population level - animals are not identified). The function tests habitat selection with the Khi2 of Pearson and log-likelihood Khi2 (recommended, see Manly et al. 2003). The Manly selectivity measure (selection ratio = used/available) is computed, the preference / avoidance is tested for each habitat, and the differences between selection ratios are computed and tested.

`widesII` computes the selection ratios with design II data (same availability for all animals, but use measured for each one). Tests of identical habitat use for all animals, and of habitat selection are also provided.

`widesIII` computes the selection ratios for design III data (when the use and the availability are measured for each animal - see examples on the wild boar below). Habitat selection is tested using a Chi-square for each animal, and the overall habitat selection is also tested.

Value

`widesI` returns a list of the class `wiI`. `widesII` returns a list of class `wiII`. `widesIII` returns a list of class `wiIII`. These objects are all inheriting from the class `wi`. They have the following components:

`used.prop`	the proportion of use for each resource type.
`avail.prop`	the proportion of available resources.
`wi`	the Manly selectivity measure (selection ratio: used/available).
`se.wi`	the standard error of the selection ratios.
`comparisons`	a list with the following components: **`diffwi`** a matrix with the differences of the selection ratios for each pair of resource type. **`ICdiffupper`** a matrix containing the upper limit of confidence interval on the differences of the selection ratios between each pair of resource type. **`ICdifflower`** a matrix containing the lower limit of confidence interval on the differences of the selection ratios between each pair of resource type. **`signif`** the ranking matrix, with the sign of the differences between the resource type in row and the resource type in column. When the difference is significant, the sign is tripled.
`profile`	the profile of preferences: resource types are sorted so that the left type is the most preferred and the right type is the most avoided. Habitats for which the selection ratios are not significant are connected by a line.
`alpha`	the parameter `alpha` of this function.
`avknown`	the parameter `avknown` of this function.
`se.used`	only for designs I, the standard error of the proportion of use.

`se.avail`	only for designs I, the standard error of the available proportion.
`chisquwi`	only for designs I, the results of Chi-Square tests of the hypothesis that the selection ratios are in average equals to zero.
`Bi`	only for designs I, equals to `wi/sum(wi)`.
`Khi2P`	only for designs I, test of random resource use (Pearson statistic).
`Khi2L`	only for designs I, test of random resource use (Log-likelihood statistic).
`Khi2L1`	only for designs II, test of identical use of habitat by all animals.
`Khi2L2`	only for designs II, test of overall habitat selection.
`Khi2L2MinusL1`	only for designs II, test of hypothesis that animals are on average using resources in proportion to availability, irrespective of whether they are the same or not (= Khi2L2 - Khi2L1).
`wij`	only for designs II and III, a matrix with the selection ratios for all animals and all resource categories.
`ICwiupper`	only for designs II and III, the upper limit of the confidence intervals on the selection ratios.
`ICwilower`	only for designs II and III, the lower limit of the confidence intervals on the selection ratios.
`Khi2Lj`	only for designs III, the test of habitat selection for each animal.
`Khi2L`	only for designs III, the global test of overall habitat selection.

Author(s)

Clément Calenge ⟨calenge@biomserv.univ-lyon1.fr⟩

References

Manly B.F.J., McDonald L.L., Thomas, D.L., McDonald, T.L. & Erickson, W.P. (2003) *Resource selection by animals - Statistical design and Analysis for field studies. Second edition* London: Kluwer academic publishers.

Thomas D. L. and Taylor E. J. (1990) Study designs and tests for comparing resource use and availability. *Journal of Wildlife Management*, **54**, 322–330.

See Also

`compana` for another analysis with this kind of data.

Examples

```
############################
## Example of moose (Manly et al., 2003, p.52)
## Known available proportions on design I data
moose.avail <- c(0.34, 0.101, 0.104, 0.455)
moose.used <- c(25, 22, 30, 40)
names(moose.used) <- c("InBurnInterior",
                       "InBurnEdge",
                       "OutOfBurnEdge",
                       "OutOfBurnFurther")
names(moose.avail) <- names(moose.used)
```

```
## Computation of wi
(wiRatio <- widesI(moose.used, moose.avail))

## plot the values of the selection ratios
opar <- par(mfrow=c(2,2))
plot(wiRatio)

par(opar)



############################
## Example of Elk (Manly et al., 2003, p.62)
## Estimated available proportions on design I data
elk.avail <- c(15, 61, 84, 40)
elk.used <- c(3, 90, 181, 51)
names(elk.used) <- c("0%", "1-25%", "26-75%", ">75%")
names(elk.avail) <- names(elk.used)
## Computation of wi
(wiRatio <- widesI(elk.used, elk.avail, avknown=FALSE))

## plot the values of the selection ratios
opar <- par(mfrow=c(2,2))
plot(wiRatio)

par(opar)



############################
## Example of Bighorn (Manly et al., 2003, p.67)
## Known available proportions on design II data
data(bighorn)
## Computation of wi
(wi <- widesII(bighorn$used, bighorn$availT, alpha = 0.1))

## plot the values of the selection ratios
opar <- par(mfrow=c(2,2))
plot(wi)


############################
## Example of Bighorn (Manly et al., 2003, p.74)
## Estimated available proportions on design II data
## Computation of wi
(wi <- widesII(bighorn$used, bighorn$availE, avknown = FALSE, alpha = 0.1))

## plot the values of the selection ratios
plot(wi)

par(opar)

############################
## Example of Wild boar
## Estimated available proportions on design III data
data(puechdesIII)
used <- puechdesIII$used
```

```
available <- puechdesIII$available

## calculation of the selectio ratios
## with sampled availability
(i <- widesIII(used,available, avknown = FALSE, alpha = 0.1))

opar <- par(mfrow = c(2,2))
plot(i)

par(opar)
```

Index