Exploring Habitat Selection by Wildlife with adehabitat

Clément Calenge

Université de Lyon

Abstract

Knowledge of the environmental features affecting habitat selection by animals is important for designing wildlife management and conservation policies. The package **adehabitat** for the R software is designed to provide a computing environment for the analysis and modelling of such relationships. This paper focuses on the preliminary steps of data exploration and analysis, performed prior to a more formal modelling of habitat selection. In this context, I illustrate the use of a factorial analysis, the K-select analysis. This method is a factorial decomposition of marginality, one measure of habitat selection. This method was chosen to present the package because it illustrates clearly many of its features (home range estimation, spatial analyses, graphical possibilities, etc.). I strongly stress the powerful capabilities of factorial methods for data analysis, using as an example the analysis of habitat selection by the wild boar (*Sus scrofa* L.) in a Mediterranean environment.

Keywords: data analysis, ecology, exploration, habitat selection, K-select analysis, radio-tracking, R.

1. Introduction

Habitat selection by wildlife is an important aspect of ecology. Studies of habitat selection provide information on environmental characteristics needed by animals, essential knowledge for the development of wildlife management and conservation policies. The habitat of an organism is often described by only one categorical variable (e.g., vegetation type), and the study of habitat selection by a species often consists of the identification of the preferred or avoided habitat types (e.g., Neu *et al.* 1974). This is mainly carried out by comparing the availability of each habitat type (e.g., measured by its percent cover in the study area) to the use of that habitat type by the species (e.g., measured by the density of animals).

Habitat is defined as the "conditions and resources present in an area that produce occupancy (including survival and reproduction) by an organism" (Hall *et al.* 1997). In essence, habitat is a multivariate concept. Many authors have stressed the need to take this into account in habitat selection studies (e.g., Morrison *et al.* 1992). Fortunately, the recent increase in availability of Geographic Information Systems (GIS) has facilitated the studies of several environmental variables at a time (e.g., elevation, slope, etc.), in habitat selection.

At the same time, many statistical methods have been developed to model habitat selection. For example, the resource selection functions (Manly *et al.* 2002), model the probability that an organism will use a resource unit (e.g., a pixel of a raster map) as a combination of the environmental variables measured on it. These methods are widely used, as they rely on the generalized linear model, which is available in most standard statistical packages.

Most of these modelling methods suppose that "the modeler knows the limiting factors that influence the distribution and abundance of the study organism" (Boyce and McDonald 1999), an important issue in most studies. Thus, according to Guisan and Zimmermann (2000), "One of the most difficult tasks is to decide which explanatory variables, or combination of variables, should enter the model". Actually, several authors defend the idea that the modelling step should be preceded by an exploration step (hindcasting, see Morrison *et al.* 1992; Calenge *et al.* 2005). It is necessary to first explore the data using statistical methods suitable for this kind of exploration (e.g., factorial analyses), in a flexible computing environment. The package **adehabitat** for the R software (R Development Core Team 2007) is intended for the exploration and modelling of the interactions between vegetal or animal species and their environment (Calenge 2006).

As previously noted, the study of habitat selection by an organism is generally carried out by comparing its habitat use to habitat availability. Thomas and Taylor (1990) have distinguished three kinds of designs, according to the levels at which use and availability are measured. With designs of type I, the individual animals are not identified (e.g., a census of the population results in a point pattern, where each point corresponds to an individual), and both use and availability are measured at the scale of the population. Designs of type II rely on identified individuals (e.g., using radio-tracking) and the use is measured for each one, but availability is still measured at the scale of the population (e.g., a study of the distribution of animals' home ranges [use] on a study area [availability]). Finally, designs of type III also rely on identified individual animals, but both use and availability are measured at the scale of the individual (e.g., a study of the selection of the feeding sites [use] by the animals inside their home range [availability]).

Here, I illustrate one particular aspect of habitat selection exploration, using the K-select analysis (Calenge *et al.* 2005). This factorial analysis allows the exploration of habitat selection in studies using designs of type III (e.g., selection of the feeding sites by the individual animals within their home range). I chose to focus on designs of type III in this paper, because: (i) radio-tracking is used in most habitat selection studies (Schooley 1994), and (ii) once the analysis of type III designs is understood, it is straightforward to generalize it to type II designs (e.g., see Calenge *et al.* 2005). Furthermore, I chose to focus on K-select analysis, as the factorial analyses are not commonly used in habitat selection studies, despite

their undeniable qualities. The package **adehabitat** provides a suitable environment for this kind of analysis. However, it is important at this point to note that many other methods are available in **adehabitat** for both modelling and exploration of habitat selection, and for all kinds of study designs (these are described in section 3). In this paper, I first describe the principle of K-select analysis. I then describe the contents of the package **adehabitat**, as well as the "philosophy" underlying the package. Finally I give an example of use of the K-select analysis on data collected on the wild boar in a Mediterranean environment.

2. The K-select analysis

The K-select analysis takes place in the framework of the duality diagram (Escoufier 1987). This family of methods groups all analyses that can be considered to be extensions of principal component analysis (PCA). The duality diagram is characterized by a triplet of matrices (Dray and Dufour 2007, same volume): (1) a table to be analyzed, (2) a diagonal matrix of weights for the columns of this table, and (3) a diagonal matrix of weights for its rows. Each column of the table defines a dimension in a multidimensional space, and the rows of the table define a cloud of points in this space. The methods belonging to the framework of the duality diagram identify a restricted set of axes in this multidimensional space, so that a given statistic - the inertia - is maximized on the first axes of the analysis. This is done thanks to a non-centered PCA of the table, weighted by rows and by columns by the two weight matrices. Actually, the PCA of the table assigns scores to its rows and its columns, so that the weighted sum of squares of the row scores, which is termed "inertia" is maximized, and the successive axes returned by the analysis are two-by-two orthogonal (uncorrelated). Then, a graphical exploration of the row and column scores allow the user to draw conclusions about the biological system under study. The table to be analyzed and the weight matrices vary with the method used, which in turn changes the biological meaning of the inertia (Greenacre 1984; Doledec *et al.* 2000). Common methods, such as discriminant analysis, canonical correspondence analysis, or correspondence analysis, belong to this family.

The K-select analysis belongs to the duality diagram. It has been fully described by Calenge *et al.* (2005), so here I give only a short description. Suppose that the study area is described by a set of p raster maps of environmental variables, each covering the same area and with the same resolution. K animals are monitored using radio-tracking devices, and a home range can be computed for each one. For a given animal k, the home-range contour delimits an area with n_k available pixels.

First, consider the case of only one animal k. Let the table $\mathbf{X}_k$ ($n_k \times p$) contain the values of the p environmental variables in each one of the n_k pixels of its home range. Each column of $\mathbf{X}_k$ defines a dimension of a multidimensional space, the ecological space. Each row of this table corresponds to an "available point" in this space. We consider that each column of the matrix $\mathbf{X}_k$ is centered: in other words, we transform the table $\mathbf{X}_k$ so that the origin of the ecological space (a p-vector of zeros) corresponds to the centroid of the cloud of available points.

Moreover, let $\mathbf{u}_k$ be a n_k-vector containing the utilization weights of the pixels by the animal k. They may be, for example, the proportions of animals' relocations in the n_k pixels of its home range. Note that the utilization weights should sum to one (i.e., $\mathbf{u}_k^\top \mathbf{1}_{n_k} = 1$, where $\mathbf{1}_{n_k}$ is the n_k-vector of 1s, and $\mathbf{u}_k^\top$ is the transpose of $\mathbf{u}_k$). The subset of available points in ecological space for which the utilization weights is greater than zero defines the set of "used points".

Now consider the vector $\mathbf{m}_k = \mathbf{X}_k^\top \mathbf{u}_k$. It contains the coordinates of the centroid of the cloud of used points in ecological space. However, since the table $\mathbf{X}_k$ has been centered, the origin of this space is the centroid of the cloud of available points. Therefore, the vector $\mathbf{m}_k$ connects the centroid of the cloud of available points (origin of the space) to the centroid of the cloud of used points. It is called the marginality vector, and its squared length $\mathbf{m}_k^\top \mathbf{m}_k$ corresponds to one measure of habitat selection, marginality. The larger the marginality, the greater the differences between average use and average availability, and the stronger the habitat selection.

Now, consider the case of K animals. Habitat availability is defined by the pixels included within the home range contours, and thus differs among animals. These availabilities are described by the "global" matrix $\mathbf{X}^\bullet$:

$$\mathbf{X}^\bullet = \begin{pmatrix} \mathbf{X}_1^\bullet \\ \mathbf{X}_2^\bullet \\ ... \\ \mathbf{X}_K^\bullet \end{pmatrix}$$

where the $(n_k \times p)$ matrix $\mathbf{X}_k^\bullet$ gives the value of the environmental variables in each pixel of the home range of the animal k. An important point is that all the columns of $\mathbf{X}^\bullet$ are centered and have unit variance. Thus, the origin of the space defined by the columns of any matrix $\mathbf{X}_k^\bullet$ is the centroid of the cloud of points defined by the rows of $\mathbf{X}^\bullet$ (i.e., the centroid of the clouds of points available to the animals, when all animals are pooled).

For each animal k, one extracts the corresponding table $\mathbf{X}_k^\bullet$, and center its columns to obtain the table $\mathbf{X}_k$ described above. Now, the origin of the space defined by the columns of $\mathbf{X}_k$ is the centroid of the cloud of points only available to the animal k. As this recentering is done for all animals, it is equivalent to a "shift" of the clouds of available points in the ecological space, so that the marginality vectors all have a common origin (see Calenge *et al.* 2005, and especially their fig. 1).

The marginality vectors $\mathbf{m}_k$ are then computed following the guidelines given above for one animal. Finally, the $(p \times K)$ table $\mathbf{M}$ can be computed:

$$\mathbf{M} = (\mathbf{m}_1 | \mathbf{m}_2 | ... | \mathbf{m}_K)$$

The K-select analysis is the non-centered principal component analysis (PCA) of the triplet $(\mathbf{M}, \mathbf{D}_K, \mathbf{Q})$. The matrix $\mathbf{Q}$ is a diagonal matrix $(p \times p)$ containing the weights of the rows of $\mathbf{M}$ (weights of the environmental variables). When all the environmental variables are numeric, all these weights are equal to 1. If categorical variables are included in the analysis, the weight matrix $\mathbf{Q}$ is adjusted using the procedure described by Hill and Smith (1976, see

next section). The matrix $\mathbf{D}_k$ is a diagonal $(K \times K)$ matrix of column weights d_k (by default, d_k is proportional to the number of relocations of the animal k). Note that the analysis is non-centered, which means that the table $\mathbf{M}$ is not centered before the PCA, unlike in classical PCA. This non-centering ensures that the origin of the ecological space corresponds to the centroids of available points of all animals on the factorial axes. The non-centered PCA therefore returns the directions in ecological space where the centroids of the used points are, on average, the furthest from the centroids of available points (i.e., the directions where the marginality vectors are, on average, the largest). This PCA consists of the computation of the eigenstructure of the matrix $\mathbf{V}$:

$$\mathbf{V} = \mathbf{M}^\top \mathbf{Q}\mathbf{M}\mathbf{D}_K$$

The eigenvalues λ_i of $\mathbf{V}$ measure the amount of marginality explained by the corresponding eigenvectors. The K-select analysis therefore maximizes the marginality explained by the first principal axes. The results of this non-centered PCA can be interpreted in the same way as a classical PCA. Scores can be defined for both the rows (environmental variables) and the columns (animals) of $\mathbf{M}$. Indeed, let the K-vector $\mathbf{f}_i$ be the ith eigenvector of $\mathbf{V}$. This vector contains the scores of the animals on the ith principal axis of the analysis. The scores of the environmental variables on the ith principal component are stored in the vector $\mathbf{g}_i = \mathbf{M}^\top \mathbf{Q}\mathbf{f}_i$.

Note that the marginality vectors have been recentered so that they all have a common origin. It is possible to compute the coordinates of the original (i.e., not recentered) vectors of marginality on the axes of the analysis. First, one projects the rows of the table $\mathbf{X}^\bullet$ on the axes of the analysis:

$$\mathbf{l}_i = \frac{1}{\sqrt{\lambda_i}}\mathbf{X}^\bullet \mathbf{Q}\mathbf{g}_i$$

The vector $\mathbf{l}_i$ contains the scores of the rows of $\mathbf{X}^\bullet$ on the axes of the ith axis of the analysis. From this calculation, it is possible to derive the $(n_k \times r)$ table $\mathbf{L}_k$ containing the scores of the n_k pixels available to the animal k on the first r axes of the analysis (with r chosen so that most of the marginality is explained by the retained axes). Then, for each animal k, one computes the vector $\mathbf{a}_k$, containing the coordinates of the centroid of the cloud of available points on the first r axes of the analysis:

$$\mathbf{a}_k = \mathbf{L}_k^\top \mathbf{1}_{n_k}$$

Finally, for each animal k, one computes the vector $\mathbf{s}_k$, containing the coordinates of the centroid of the cloud of used points:

$$\mathbf{s}_k = \mathbf{L}_k^\top \mathbf{u}_k$$

These two centroids can be plotted on a factorial map, and the non-centered marginality vector connects the two centroids (see example).

After choosing the number of axes to keep, some schools of thought recommend the use of a second rotation of the factorial axes in this reduced subspace, to find another set of

axes for which the meaning is clearer. For example, the varimax method rotates the chosen factorial axes to find new axes with the largest variance of the variables scores (Jollife 2004). The resulting axes are no longer those on which the marginality is maximized (though the subspace in which this rotation is performed is still the one in which the marginality is at its maximum), but they are more easily interpreted. Many choices are possible for this second rotation (See Jollife 2004), which leaves some arbitrariness or subjectivity in the interpretation of the results. Nevertheless, if the user wants to use such methods, many functions, such as `varimax` and `promax`, are available in the R environment.

3. The package adehabitat

As noted previously, **adehabitat** provides a framework for the exploration and modelling of habitat selection using the three types of study designs. Because the philosophy of **adehabitat** is mainly centered on data exploration, it strongly relies on the package **ade4** (Dray and Dufour 2007, same volume). This package contains numerous functions to perform factorial analyses (Escoufier 1987). Actually, **adehabitat** is organized around four poles: (i) the management of raster maps, (ii) home range estimation, (iii) the analysis of animals' movements, and (iv) the analysis of habitat selection / ecological niche.

First, **adehabitat** can be used to import and export raster maps from/to Geographic Information Systems (e.g., see the function `import.asc`). It provides basic spatial operations, such as the computation of buffers around points or lines (see the function `buffer`), the identification of the value of environmental variables at given spatial locations (`join.asc`), the numbering of points in each pixel of a map (`count.points`), etc. These functions allow the user to manage information about the environment of the studied species.

Moreover, **adehabitat** contains many functions that facilitate the analysis of radio-tracking data. Many methods are available for home range estimation, including the Minimum convex polygon (Mohr 1947, function `mcp`), the kernel estimator (Worton 1995, function `kernelUD`), the cluster home range (Kenward *et al.* 2001, function `clusthr`) or the nearest neighbour convex hull (Getz and Wilmers 2004, function `NNCH`). The home range provides a static point of view on the use of space by animals. A more dynamic analysis is possible with **adehabitat**, as several functions allow the analysis of animals movements (see the help page of `as.ltraj` and Calenge *et al.* in prep.). All these functions only consider the movements of the animals, without any reference to environment factors.

Finally, several functions provide statistical methods for analysis of habitat selection by wildlife. Many of them are factor analyses of the niche or of habitat selection, but other methods are also available. For studies of design type I, they include the selection ratios (Manly *et al.* 2002, function `wi`), the ecological niche factor analysis (Hirzel *et al.* 2002, function `enfa`), the Mahalanobis distances (Clark *et al.* 1993, function `mahasuhab`) and their factorial partitioning (Calenge *et al.* 2004, function `madifa`), or the algorithm DOMAIN (Carpenter *et al.* 1993, function `domain`). Other common methods, such as resource selection functions (Manly *et al.* 2002) can also be used by combining the functions of **adehabitat** with the rest of the R environment. For design type II and III studies, the package includes functions imple-

menting the compositional analysis (Aebischer and Robertson 1992, function `compana`), the eigenanalysis of selection ratios (Calenge and Dufour 2006, function `eisera`) or the K-select analysis (Calenge *et al.* 2005, function `kselect`).

I now describe more specifically the functions and object classes available to perform a K-select analysis. The K-select analysis implies several spatial operations. One must estimate the home range of animals, count the number of relocations in each pixel of a raster map, and identify the pixels of the map inside the home-range limits. As noted previously, **adehabitat** deals mostly with raster maps (i.e., maps where the spatial information is coded by a matrix of pixels). A raster map is stored as a matrix of class "`asc`", in which each cell corresponds to one pixel of the map. The most common way to create a raster is to import an Arcview ASCII raster file with the function `import.asc`. This file format can be generated by most Geographic Information Systems (Arcview, GRASS, Landserf, etc.).

As the K-select analysis is a multivariate analysis, one has to deal with a list of several maps of the same study area, all with the same resolution. It is convenient to convert this list of maps into a multi-layered map of class "`kasc`". A map of class "`kasc`" is a data frame giving the values of the environmental variables (columns) in each pixel of the map (rows). It can be created with the function `as.kasc`, which takes as its argument a list of named objects of class "`asc`".

The aim of the analysis is to identify the variables affecting site selection by animals within their home range. Therefore, one must first estimate the home-range contour. As noted above, many home range estimation methods are available in **adehabitat**. Alternatively, one can also define the area available to the animals thanks to a buffer around the relocations, using the function `buffer.ani`.

The identification of pixels located inside the home range contours depends on the method chosen for home-range estimation. Here, we estimate the minimum convex polygon with the function `mcp`. This function returns a data.frame of class "`area`", with three columns. The first column contains the ID of the animal, and the second and the third columns contain the coordinates of the vertices of the home range. These home ranges should be rasterized using the function `hr.rast`. The result of this function is again an object of class "`kasc`", with the same number of pixels (rows) as the map of the study area, and with one column per animal. Each pixel of a given map takes a value of 1 if the pixel is inside the home range or `NA` (missing value) if it is not.

The last component needed by the K-select analysis is the multi-layer map of class "`kasc`" containing the number of relocations in each pixel of the map (rows) for each animal (columns). This map can be computed using the function `count.points.id`.

Three main objects result from the above operations: (i) a multi-layer map of the environmental variables in the study area, (ii) a multi-layer map of the home ranges of the animals, and (iii) a multi-layer map of the number of relocations in each pixel of the study area. These three maps are then combined into a list of class `sahrlocs`, using the function `as.sahrlocs`.

Several graphical methods are available in **adehabitat** to explore habitat selection with this class of objects. Among other things, objects of this class can be used to derive the arguments of the function `kselect`, using the function `sahrlocs2kselect`. This function returns a list with three components. The component named `tab` is a data frame containing the table $\mathbf{X}_u^\bullet$. It is similar to the table $\mathbf{X}^\bullet$ described in the previous section, except that the columns of this table are neither centered nor scaled to unit variance. The component named `weight` is the vector containing the concatenated vectors of utilization weights (i.e., the vector $(\mathbf{u_1}^\top, \mathbf{u}_2^\top, ..., \mathbf{u}_K^\top)^\top$). The component named `factor` contains a factor which can be used to split the previous components into K matrices $\mathbf{X}_{uk}^\bullet$ and K vectors $\mathbf{u}_k$, respectively.

At this point, note that the table $\mathbf{X}_u^\bullet$ needs to be transformed into the table $\mathbf{X}^\bullet$ described in the previous section (i.e., with each column centered and having a unit variance). This is achieved thanks to a PCA of the component `tab` of the list returned by `sahrlocs2kselect`. This PCA is performed using the function `dudi.pca` of the package **ade4** (dudi stands for "duality diagram"). The object returned by this function is a list of class `dudi` containing, among other things, the centered table needed for the K-select analysis. The component `tab` of the resulting object of class `dudi` is the matrix $\mathbf{X}^\bullet$ described in the previous section.

Note that the preliminary use of a simple factorial analysis to suitably transform the table is standard practice in the packages **ade4** and **adehabitat** (e.g., see the functions `enfa`, `madifa`, `discrimin` or `niche`). Although this factorial analysis, in itself, is not necessarily of interest, this is an easy way to transform the table and to define suitable weights for its rows and columns.

Although we will mainly focus on numeric environmental variables (e.g., elevation), it is important to note that categorical variables (e.g., vegetation type) may also be included in the analysis, using the procedure described by Hill and Smith (1976). The categories are coded by a set of dummy variables (zero indicating the absence of a category, and one, its presence) prior to the analysis. For a given categorical variable, the weights of the dummy variables coding for the categories (in the matrix $\mathbf{Q}$) are computed so that the more frequent categories have more weight, and so that the sum of the weights of the dummy variables is equal to one. All the variables of the analysis therefore have the same weight, whether numeric or categorical. Thus, if the table to be analyzed included categorical variables, the use of `dudi.pca` prior to the use of `kselect` should be replaced by `dudi.mix`, to compute suitable weights for the levels of the categorical variables.

Finally, the function `kselect` can be applied, taking as arguments the results of the functions `dudi.pca` and `sahrlocs2kselect`. The results of the analyses can be explored using the graphical functions `plot.kselect` and `kplot.kselect`.

4. Example

Here, we use the dataset `puech` of the package **adehabitat**. This dataset is a subset of the data on wild boar collected at Puechabon (South of France) by D. Maillard during his Ph.D (Maillard 1996). More information can be obtained on the help page of this dataset, by typing

```
?puech.
```

In this section, I describe the steps to perform a K-select analysis of these data. First, we load the package **adehabitat** and the dataset:

```
R> library("adehabitat")
R> data("puech")
R> names(puech)
```

```
[1] "relocations" "maps"
```

This data set is a list with two components. The component named `relocations` is a data frame containing the relocations of the diurnal resting sites of six wild boars, with 30 relocations per animal:

```
R> locs <- puech$relocations
R> head(locs, 3)
```

```
   name      x       y
1 Brock 698811 3160911
2 Brock 699809 3161496
3 Brock 699474 3161193
```

```
R> table(locs[, 1])
```

```
  Brock    Chou    Jean   Kinou Schnock Suzanne
     30      30      30      30      30      30
```

The component named `map` is a list with 9 named elements: each element is a map of class "`asc`" (Figure 1). Because all the maps cover the same area with the same resolution, they can be combined into a multi-layer map of class "`kasc`".

```
R> kasc <- as.kasc(puech$maps)
```

The maps can be displayed by typing (Figure 1):

```
R> image(kasc)
```

These maps can also be explored interactively using `explore.kasc`. Note that it is possible to convert this object into a data frame with the function `kasc2df`, and thatfactorial analyses can be used to explore this data frame (e.g., PCA with the function `dudi.pca` from **ade4**).

We then estimate the home range of the animals with the minimum convex polygon. The function `mcp` takes as an argument a data frame with two columns containing the coordinates of the relocations, and a vector containing the animals IDs for each one:
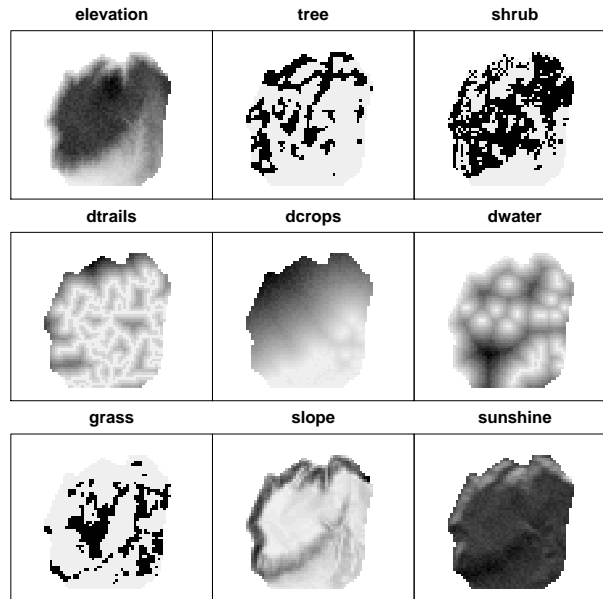
Figure 1: The raster maps of the study area at Puéchabon, South of France. The studied variables are (left to right, top to bottom) elevation, tree cover, shrub cover, distance to recreational trails, distance to crops, distance to water points, grass cover, slope and sunshine. The function `image.kasc` used to draw this map gives a general view of the spatial distribution of the variables in the study area. If the user requires the precise values of the variables, then the functions `explore.kasc` or `plot.asc` may be used.

```
R> pc <- mcp(locs[, 2:3], locs[, 1], percent = 100)
```

The parameter `percent` is equal to 100 minus the percentage of relocations excluded from the computation. The closest relocations to the centroid of the distribution of relocations are kept to compute the limits of the home ranges. These polygons can be plotted on the map of the area using the functions `image.asc` and `plot.area` (Figure 2):

```
R> par(mar = c(0, 0, 0, 0))
R> image(puech$maps$elevation, axes = FALSE)
R> plot(pc, add = TRUE, lwd = 3, colpol = "grey")
```

The function `par` allows for control of the margin size. The argument `axes` of the function `image` is set to `FALSE` so that no axes are displayed on the graph. The argument `add` of the function `plot` is used to add the polygons to the previous plot. The argument `lwd` controls the line width and the argument `colpol` controls the color of the polygons.

We then rasterize these home ranges, using the function `hr.rast`. This function takes as an argument the object of class "`area`" returned by `mcp`, as well as a raster map:

```
R> pcr <- hr.rast(pc, kasc)
```
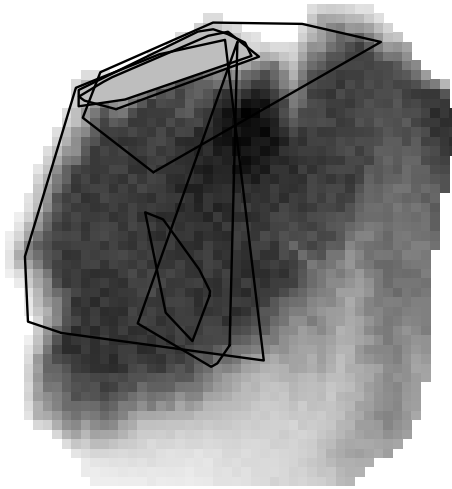
Figure 2: Home ranges of the wild boars monitored using radio-tracking at Puéchabon, displayed on an elevation map.

Again, note that the result can be displayed using `image.kasc`. We then count the number of relocations in each pixel of the map `kasc`. The function `count.points.id` takes as an argument a data frame with two columns giving the coordinates of the relocations, a vector indicating the animals ID, and a raster map:

```
R> cp <- count.points.id(locs[, 2:3], locs[, 1], kasc)
```

Then, the three objects `cp`, `pcr` and `kasc` can be combined into an object of class `sahrlocs`.

```
R> sahr <- as.sahrlocs(cp, pcr, kasc)
```

The functions `plot.sahrlocs` and `image.sahrlocs` provide additional methods to explore these data. We derive the data structure needed for the K-select analysis:

```
R> jj <- sahrlocs2kselect(sahr)
R> names(jj)
```

```
[1] "tab"    "factor" "weight"
```

I described the components of this list in the previous section. We then perform the preliminary PCA of the component `tab` of the object `jj`.

```
R> prico <- dudi.pca(jj$tab, scannf = FALSE)
```

The argument `scannf`, when set to `TRUE`, allows the user to interactively select the number of principal components to keep in the PCA (we do not need it here, as this PCA is performed only to transform the table). Finally, we perform the K-select analysis:

```
R> ks <- kselect(prico, jj$factor, jj$weight, scannf = FALSE)
R> ks


Duality diagramm
class: kselect dudi
$call: kselect(dudi = prico, factor = jj$factor, weight = jj$weight,
    scannf = FALSE)


$nf: 2 axis-components saved
$rank: 6
eigen values: 1.797 0.4353 0.1379 0.05759 0.02727 ...
  vector    length mode    content
1 $cw       6      numeric column weights
2 $lw       9      numeric row weights
3 $eig      6      numeric eigen values
4 $initfac  1200   numeric initial factor
5 $initwei  1200   numeric row weights of inittab


  data.frame nrow ncol content
1  $tab      9    6    modified array
2  $li       9    2    row coordinates
3  $l1       9    2    row normed scores
4  $co       6    2    column coordinates
5  $c1       6    2    column normed scores
6  $initab   1200 9    initial table centered per animal
7  $as       2    2    axis upon kselect axis
8  $ls       1200 2    rows of initab upon kselect axis
9  $mus      6    2    mean use on kselect axis
10 $mav      6    2    mean available on kselect axis
```

The function `plot.kselect` displays a summary of the results (see the help page of the function, Figure 3):

```
R> plot(ks)
```

Note that the function `kplot.kselect` can also be used to draw more complete factorial maps of the ecological space.

The eigenvalues of this analysis indicate that two axes are sufficient to explain most of the marginality of the data set, though the first one is much larger than the second. Actually, the first axis opposes the areas close to the gorges of the Hérault river on the positive side (low elevation, high slope, low sunshine, far from crops and trails and close to the water),
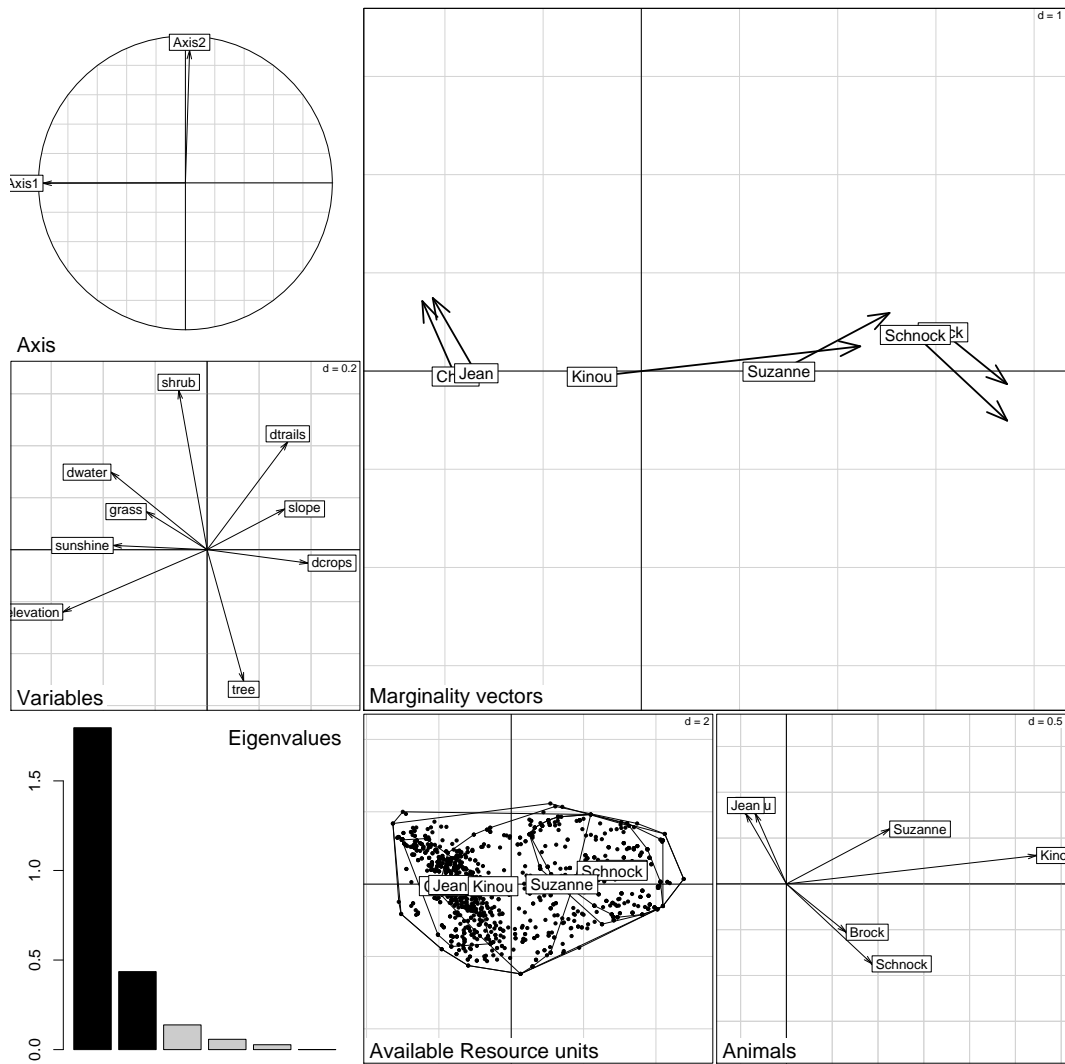
Figure 3: Results of the K-select analysis carried out on the radio-tracking data collected on the wild boars of Puéchabon (South of france). The graphs are (top to bottom, left to right): (i) the correlations between the axes of the K-select analysis and the first two axes of the PCA of the table $\mathbf{X}^\bullet$, (ii) the scores of the environmental variables on the axes of the K-select analysis, (iii) the eigenvalues of the analysis, (iv) the uncentered vectors of marginality (the origin of the arrow indicates the centroid of the cloud of available points and its end indicates the centroid of the cloud of used points), (iv) scores of the available resource units on the axes of the analysis and (v) scores of the recentered marginality vectors on the first axes of the analysis.

to the plateau of Puéchabon on the negative side (opposite characteristics). The second axis opposes the areas with a dense shrub cover on the positive side to the areas with a low cover on the negative side.

The uncentered marginality vectors (main graph on Figure 3) reveal that the animals mostly

select the gorges of the Hérault river, but that great variation occurs among animals. Indeed, when the gorges are available in the home ranges of the animals (i.e., when the basis of the marginality vector is close to the positive side of the first axis), they are strongly selected (the marginality vectors point to the positive side). When the gorges are not available (animals "Chou" and "Jean"), the animals rather select a dense shrub cover (the marginality vectors point at the positive side of the second axis). These results therefore indicate a functional response of the wild boar (i.e., a habitat selection changing according to habitat availability). As indicated by the results of the K-select analysis, this functional response occurs on two habitat variables: the gradient gorges-plateau (first axis of the K-select analysis) and the shrub cover (second axis of the analysis).

Based on these results, we may draw a graph that shows this functional response more clearly. We focus on the two variables on which this response occurs: the gorges-plateau gradient and the shrub cover. First, we define two habitat types: the gorges and the plateau. The gorges are defined as the areas with an elevation lower than 200 metres and the plateau corresponds to the areas with an elevation above 200 metres:

```
R> hab <- jj$tab$elevation < 200
```

Note that the crops found in the southern portion of the study area, and for which the elevation is lower than 200 metres, are not located in the gorges (Figure 1). However, these crops are not included in the home range of the monitored animals (Figure 3). Therefore, from the point of view of the animals, all the habitat types located at an elevation below 200 metres are part of the gorges, so that the above definition of the gorges is still correct.

Then, we compute the available proportion of gorges in the home range of each animal:

```
R> pa <- table(jj$factor, hab)[, 2]/table(jj$factor)
R> class(pa) <- NULL
R> pa


  Brock    Chou    Jean   Kinou Schnock Suzanne
0.78947 0.00000 0.01190 0.13154 0.70000 0.36515
```

Similarly, we compute the proportion of relocations in the gorges:

```
R> sp1 <- split(hab, jj$factor)
R> sp2 <- split(jj$weight, jj$factor)
R> pu <- unlist(lapply(1:length(sp1), function(x) sum(sp2[[x]] *
+     sp1[[x]])/sum(sp2[[x]])))
R> pu


[1] 1.0000 0.0000 0.0400 0.7308 1.0000 0.8462
```

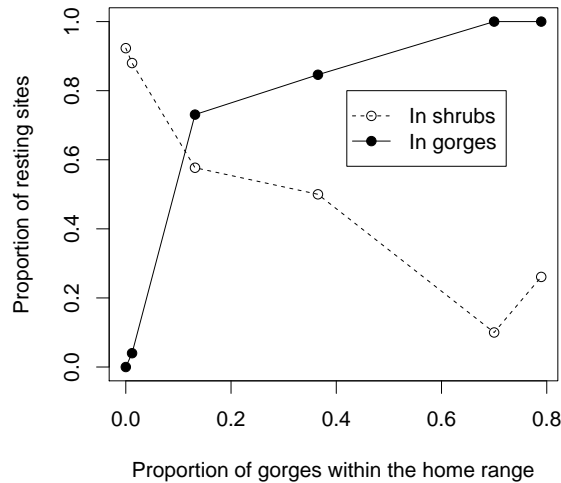We also compute the average shrub cover used by the animals:

Figure 4: Functional response of the wild boar: proportion of relocations in the gorges of the Hérault river and the shrubs, as a function of the proportion of the home range covered by the gorges, for all animals.

```
R> sp3 <- split(jj$tab$shrub, jj$factor)
R> pushrub <- unlist(lapply(1:length(sp3), function(x) sum(sp2[[x]] *
+     sp3[[x]])/sum(sp2[[x]])))
R> pushrub
```

```
[1] 0.2609 0.9231 0.8800 0.5769 0.1000 0.5000
```

And finally, we can compute the relationship between the proportion of resting sites in both the gorges and the shrubs, and the proportion of the home-range covered by the gorges (Figure 4):

```
R> plot(pa, pu, xlab = "Proportion of gorges within the home range",
+     ylab = "Proportion of resting sites", pch = 16)
R> lines(pa[order(pa)], pu[order(pa)])
R> points(pa[order(pa)], pushrub[order(pa)])
R> lines(pa[order(pa)], pushrub[order(pa)], lty = 2)
R> legend(0.42, 0.8, legend = c("In shrubs", "In gorges"), pch = c(1,
+     16), lty = c(2, 1))
```

A "constant" habitat selection would have resulted in linear relationships between use and availability (Mysterud and Ims 1998). Instead, these relationships are curved, indicating a functional response (i.e., a selection depending on habitat availability). This illustrates the power of factorial analyses, as such responses are very difficult to identify in ecology (Mysterud and Ims 1998). Although we only have six animals, the results are strikingly clear.

5. Discussion

I illustrate the use of the package **adehabitat** in habitat selection studies, and more specifically, the K-select analysis within the framework of data exploration. An important aspect of this example is that of this method implies several preliminary steps (home range estimation, relocation numbering, creation of an object of class "`sahrlocs`", etc.). All of these steps could have been combined into a single function, to facilitate the application of the analysis. However, this would have been contrary to the philosophy of **adehabitat** (and more generally to the philosophy of R). At each step of the process, a new class of objects is created, which corresponds to a new view of the problem. The home-range estimation gives information about the space used by the animal, the object of class `sahrlocs` can be used for graphical exploration, etc. I mentioned some of these related aspects, and I encourage the reader to test the possibilities of the package.

The data exploration allows the user to build a conceptual model of the system under study. Then, RSF or similar methods can be used for a more formal modelling. The package `adehabitat` provides a suitable environment for this kind of analysis, as it contains a large number of functions that perform the spatial operations needed by this methodology (spatial joining, buffering, etc.). Thus, in the example presented above, a possible additional step would have been to model the probability of selection of a pixel by a wild boar according to two variables, one indicating the proximity of the gorges and one measuring the density of the shrub cover. I did not present this aspect, as this is beyond the scope of this paper.

Note that the map storage is quite memory expensive in **adehabitat**. Most R packages dealing with maps have this problem, although the package **sp** allows for more efficient storage for certain maps. Note that some functions in **adehabitat** provide an interface between the spatial classes of **sp** and those of **adehabitat** (see the help page of the function `spixdf2kasc`).

Despite this minor drawback, `adehabitat` has been used successfully in several fields of ecology, including population ecology (e.g., Calenge *et al.* 2005) or biogeography (e.g., Spichiger *et al.* 2004). In landscape ecology, the package **lsmetrics**, which relies strongly on **adehabitat**, has been developed to provide landscape connectivity indices (Ribeiro *et al.* 2007). **adehabitat** is still actively being developed, and new tools are regularly added, based either on contributions of users (e.g., `NNCH` or `enfa`) or on users suggestions (e.g., `kernelbb`). In addition, a tutorial explaining the spatial classes and basic spatial functions is available in the package. At the time of writing, the package contains more than 170 functions dedicated to the analysis and modelling of relationships between one or several organisms and their environment, with a special focus on factorial analyses.

Acknowledgments

would like to acknowledge the numerous helpful comments and suggestions of the users.

References

Aebischer NJ, Robertson PA (1992). "Practical Aspects of Compositional Analysis as Applied to Pheasant Habitat Utilization." In IG Priede, SM Swift (eds.), "Wildlife Telemetry: Remote Monitoring and Tracking of Animals," pp. 285–293. Ellis Horwood, London.

Boyce MS, McDonald LL (1999). "Relating Populations to Habitats Using Resource Selection Functions." *Trends in Ecology and Evolution*, **14**, 268–272.

Calenge C (2006). "The Package **adehabitat** for the R Software: A Tool for the Analysis of Space and Habitat Use by Animals." *Ecological Modelling*, **197**, 516–519.

Calenge C, Darmon G, Basille M, Loison A, Jullien JM (2004). "The Factorial Decomposition of the Mahalanobis Distances in Niche Analysis." *Ecology*. In revision.

Calenge C, Dufour AB (2006). "Eigenanalysis of Selection Ratios from Animal Radio-Tracking Data." *Ecology*, **87**, 2349–2355.

Calenge C, Dufour AB, Maillard D (2005). "K-select Analysis: A New Method to Analyse Habitat Selection in Radio-Tracking Studies." *Ecological Modelling*, **186**, 143–153.

Carpenter G, Gillison AN, Winter J (1993). "DOMAIN: a Flexible Modelling Procedure for Mapping Potential Distributions of Plants and Animals." *Biodiversity and Conservation*, **2**, 667–680.

Clark JD, Dunn JE, Smith KG (1993). "A Multivariate Model of Female Black Bear Habitat Use for a Geographic Information System." *Journal of Wildlife Management*, **57**, 519–526.

Doledec S, Chessel D, Gimaret-Carpentier C (2000). "Niche Separation in Community Analysis: A New Method." *Ecology*, **81**, 2914–2927.

Dray S, Dufour AB (2007). "The **ade4** Package: Implementing the Duality Diagram for Ecologists." *Journal of Statistical Software*, **22**(4). URL http://www.jstatsoft.org/v21/i04/.

Escoufier Y (1987). "The Duality Diagram: A Means of Better Practical Applications." In P Legendre, L Legendre (eds.), "Development in Numerical Ecology," Series G, pp. 139–156. Springer-Verlag, Berlin.

Getz WM, Wilmers CC (2004). "A Local Nearest-Neighbor Convex-Hull Construction of Home Ranges and Utilization Distributions." *Ecography*, **27**, 489–505.

Greenacre MJ (1984). *Theory and Applications of Correspondence Analysis*. Academic Press, London, UK.

Guisan A, Zimmermann NE (2000). "Predictive Habitat Distribution Models in Ecology." *Ecological Modelling*, **135**, 147–186.

Hall LS, Krausman PR, Morrison ML (1997). "The Habitat Concept and a Plea for Standard Terminology." *Wildlife Society Bulletin*, **25**, 173–182.

Hill MO, Smith AJE (1976). "Principal Component Analysis of Taxonomic Data With Multi-State Discrete Characters." *Taxon*, **25**, 249–255.

Hirzel AH, Hausser J, Chessel D, Perrin N (2002). "Ecological-Niche Factor Analysis: How to Compute Habitat Suitability Maps Without Absence Data?" *Ecology*, **83**, 2027–2036.

Jollife IT (2004). *Principal Component Analysis.* Springer-Verlag, New York, 2nd edition.

Kenward RE, Clarke RT, Hodder KH, Walls SS (2001). "Density and Linkage Estimators of Home Range: Nearest Neighbor Clustering Defines Multinuclear Cores." *Ecology*, **82**(7), 1905–1920.

Maillard D (1996). *Occupation et utilisation de la garrigue et du vignoble méditerranéens par le sanglier (Sus scrofa L.).* Phd thesis, Université de droit, d'économie et des sciences d'Aix-Marseille III.

Manly BFJ, McDonald LL, Thomas DL, MacDonald TL, Erickson WP (2002). *Resource Selection by Animals. Statistical Design and Analysis for Field Studies.* Kluwer Academic Publisher, London, 2nd edition.

Mohr CO (1947). "Table of Equivalent Populations of North American Small Mammals." *American Midland Naturalist*, **37**, 223–249.

Morrison ML, Marcot BG, Mannan RW (1992). *Wildlife-Habitat Relationships. Concepts and Applications.* The University of Wisconsin Press.

Mysterud A, Ims RA (1998). "Functional Responses in Habitat Use: Availability Influences Relative Use in Trade-Off Situations." *Ecology*, **79**, 1435–1441.

Neu CW, Byers CR, Peek JM (1974). "A Technique for Analysis of Utilization-Availability Data." *Journal of Wildlife Management*, **38**, 541–545.

R Development Core Team (2007). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Ribeiro MC, Metzger JP, Calenge C, Martensen AC, Linhares C, Lima A, Ciocheti G (2007). "The Package **lsmetrics** for Computing Landscape Connectivity Indices in R Software." *Environmental Modelling & Software.* Forthcoming.

Schooley RL (1994). "Annual Variation in Habitat Selection: Patterns Concealed by Pooled Data." *Journal of Wildlife Management*, **58**, 367–374.

Spichiger R, Calenge C, Bise B (2004). "The Geographical Zonation in the Neotropics of Tree Species Characteristic of the Paraguay-Paraná Basin." *Journal of Biogeography*, **31**, 1489 – 1501.

Thomas DL, Taylor EJ (1990). "Study Designs and Tests for Comparing Resource Use and Availability." *Journal of Wildlife Management*, **54**, 322–330.

Worton BJ (1995). "Using Monte Carlo Simulation to Evaluate Kernel-based Home Range Estimators." *Journal of Wildlife Management*, **59**, 794–800.

Affiliation:

Clément Calenge
Université de Lyon
Laboratoire de Biométrie et Biologie Evolutive (UMR 5558); CNRS
Université de Lyon
université Lyon 1
43, Boulevard du 11 Novembre 1918
F-69622 Villeurbanne Cedex, France
Office national de la chasse et de la faune sauvage
95 rue Pierre Flourens
F-34000 Montpellier, France
Telephone: +33/4/72432757
E-mail: calenge@biomserv.univ-lyon1.fr