Coupling Principal Component Analysis and GIS to map deer habitats

Authors: Nathalie Pettorelli, Stephane Dray, and Daniel Maillard
Source: Wildlife Biology, 11(4) : 363-370
Published By: Nordic Board for Wildlife Research
Coupling Principal Component Analysis and GIS to map deer habitats

Nathalie Pettorelli, Stephane Dray & Daniel Maillard


We aimed to define at a relevant scale the spatial pattern of major vegetation types available to deer in order to characterise habitat quality variations within our population area. We analysed data on the timber stand and the shrub layer collected in 1993 in the 2,614 ha Chizé reserve in western France. Multidimensional analyses (Principal Component Analysis and biplot) and a Geographic Information System (GIS) were used to extract most of the variation in vegetation data collected at the 4-ha resolution. At the timber stand level, two vegetation types occurred within the reserve: an oak Quercus sp. stand in the north, and a beech Fagus sylvatica stand in the south. This classification accounted for 29.6% of the total variability of the timber stand data base. At the shrub layer scale, three vegetation types were distinguished: hornbeam Carpinus betulus dominated coppices in the northeast part of the oak stand, maple Acer sp. dominated coppices in the northwest part of the oak stand, and no shrub layer in the beech stand in the south. This classification accounted for 32% of the total variability of the shrub layer data base. The coupled use of multivariate analysis and GIS allowed us to assess classification of forest habitats and appears promising for use in wildlife management and research purposes. This simple and robust tool allows users to account for site variability, and can provide satisfactory spatial representations of habitat potential at multiple scales.

**Key words:** Capreolus capreolus, Chizé reserve, habitat classification, management, multidimensional statistics

Nathalie Pettorelli* & Stephane Dray**, Unité Mixte de Recherche N°5558 « Biométrie et Biologie Evolutive », Bâtiment 711, Université Claude Bernard Lyon 1, 43 Boulevard du 11 novembre 1918, 69622 Villeurbanne Cedex, France - e-mail addresses: nathalie.pettorelli@bio.uio.no (Nathalie Pettorelli); stephane.dray@umontreal.ca (Stephane Dray)

Daniel Maillard, Office National de la Chasse et de la Faune Sauvage, 95 Rue Pierre Flourens, 34098 Montpellier Cedex 5, France - e-mail: d.maillard@oncfs.gouv.fr

Present addresses:
*Centre for Ecological and Evolutionary Synthesis, Department of Biology, University of Oslo, PO Box 1050 Blindern, N-0316 Oslo, Norway
**Department of Biological Sciences, University of Montréal, C.P. 6128, Montréal, Québec H3C 3J7, Canada

**Corresponding author:** Nathalie Pettorelli

Article accepted for publication in Game and Wildlife Science (GWS) prior to the merger between WILDLIFE BIOLOGY and GWS. Therefore, this article has not been peer-reviewed by WILDLIFE BIOLOGY Associate Editors
To assess the importance of space in mammalian population dynamics (Tilman & Kareiva 1997) or as an attribute of individual variability in population dynamics (Gaillard et al. 2000), a detailed knowledge of the spatial distribution of food resources is needed. A common way of dealing with spatial habitat descriptive data is to consider only some of the available information. It thus involves lumping the population’s habitat into a few blocks, generally based on knowledge of the animal’s diet (Coulson et al. 1997) or on assumptions of plant quality (Mysterud et al. 2001). However, as habitat assessment generally implies working with numerous variables, multivariate analyses are natural tools to deal with such data. Habitat assessment also requires precise spatial representations of variability. Geographical Information Systems (GIS) have recently been shown as a powerful tool for dealing with spatial representations. Coupling multivariate analysis and GIS appears to be a powerful tool for vegetation structure analyses. There are indeed numerous studies in landscape ecology that have coupled multivariate analyses and GIS to describe vegetation structure in particular regions (Kadmon & Danin 1997, Ohmann & Spies 1998, Kadmon & Heller 1998, Guisan et al. 1999). Nevertheless, few studies (but see Morellet 1998) have encouraged wildlife managers and foresters to utilise both tools concurrently.

In the Chizé reserve in France, habitat-specific life history traits have been reported during 25 years’ monitoring of the roe deer Capreolus capreolus population (Pettorelli et al. 2001, 2002). Based on coarse data on major tree species, two broad vegetation types have been distinguished within the managed forest: the northern oak Quercus sp. stand and the southern beech Fagus sylvatica stand. The woodland type was found to affect the repartition of the main resources used by roe deer in spring and summer (Pettorelli et al. 2001), as well as the body masses of fawns (Pettorelli et al. 2001) and adults (Pettorelli et al. 2002). However, this rather crude approach to habitat definition might be far from optimal. In particular, the two broad types did not consider the shrub layer, which is one of the most important factors affecting deer habitat selection (Mysterud et al. 1999).

In this paper, we present a method integrating multivariate analyses and GIS that allows consideration of all the descriptive variables collected in order to separate major plant communities occurring within the population habitat. We assessed spatial structure at a scale that could be related to roe deer to determine whether variables other than the major stand types could affect habitat structure and quality.

### Material and methods

Forest structure data were collected in 1993 by foresters of the Office National des Forêts (ONF) in the 2,614 ha Chizé reserve in western France (46°05’N, 0°25’W; Fig. 1). This fenced forest has been managed by foresters for >200 years. The data we are using here are thus from before the storm Lothar which touched the reserve in December 1999 (Gaillard et al. 2003). The climate is oceanic with Mediterranean influences and characterised by mild winters and hot, dry summers (Table 1). According to a geological map based on whole soil measurements (Lambertin 1992), soils tend to be clayey in the northern part, chalky and limestone soils associated with marl being the rule in the southern part of the reserve.

In France, state forests are divided into numbered plots.

<table>
<thead>
<tr>
<th>Winter</th>
<th>Summer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>Precipitation</td>
</tr>
<tr>
<td>Mean</td>
<td>6.16°C</td>
</tr>
<tr>
<td>Range</td>
<td>3.85°C - 7.87°C</td>
</tr>
</tbody>
</table>

Table 1. Climatic conditions in winter (December-February) and summer (June-August) at the Chizé reserve based on data from 1977 to 1998.
(average size: 10-15 ha) bounded by forest trails. However, the Chizé forest is managed at a finer subplot scale ($\bar{x} = 4$ ha). At this particular scale, foresters in 1993 determined the four major species at the shrub layer scale and the three major species at the timber stand scale, with their percent cover. For each subparcel, the total cover was 100% at the shrub layer and at the timber stand scales. A Geographical Information System (GIS) layer of the reserve was created from a 1:10,000 National Geographic Institute (IGN) map and a 1:10,000 map developed by foresters (that has allowed digitising of subplot limits). Areas of subplots were then derived using ARC VIEW 3.2 (Mitchell 1999). Cover was determined using 1:10,000 Infra Red Colour (IRC) aerial photographs taken in June 1993 and information gathered in the reserve by foresters.

Four oak species, i.e. *Quercus robur*, *Q. pubescens*, *Q. petraea* and *Q. cerris*, occurred in the timber stand and two, i.e. *Quercus pubescens* and *Q. robur*, in the shrub layer. Four pine species, i.e. *Pinus pinaster*, *P. sylvestris*, *P. nigra* and *P. laricio*, occurred in the timber stand. Two maple species, i.e. *Acer campestre* and *A. monspessulanum*, occurred in both strata. The species were not separated in all subplots and thus were pooled into three groups: oaks, maples and pines. Rare species of coniferous and deciduous trees were pooled into two categories. Thus, we considered 10 timber stand categories, i.e. maples, pines, oaks, hornbeam *Carpinus betulus*, beech *Fagus sylvatica*, other deciduous, other coniferous, cherry tree *Prunus avium*, cedar *Cedrus* sp. and Douglas fir *Abies douglasii*, for 629 subplots. We considered five categories, i.e. maples, hornbeam, oaks, beech and other deciduous, at the shrub layer scale for 621 subplots. Data were available for more than 97% of timber subplots ($N = 648$ in the reserve) and 96% at the shrub layer scale.

We used multivariate analyses to extract the main sources of variation in the data set. Principal Component Analysis (PCA) on a table of proportions has been introduced in general ecology in the last decades (Aitchison 1983, Ter Braak 1983). PCA and the associated graphical representation based on biplot theory (Gabriel 1971, De Crespin de Billy et al. 2000) are very suitable for data in a table of proportions. On the biplot representation, cover-type categories are positioned relative to the principal axes of unit variance ($\bar{x} = 0$) and represented by arrows. Subplots are plotted according to an averaging procedure based on proportion data. This averaging procedure allows the user to display mean composition of subplots which differentiates this analysis from the usual PCA procedure (Hotelling 1933). Positions of cover-type categories on the biplot representation depend on their relative abundance and variability of vegetation composition among subplots. Rare cover-type categories are located around the origin whereas major cover-type categories are more distant. Thus, ordination of subplots allows direct examination of their vegetation composition and illustrates the main structure of the reserve. Subplots close to the origin contain all major cover-type categories or are characterised by rare cover-type categories, whereas subplots near the end of an arrow are characterised by the cover-type category the arrow represents.

Because the species considered varied at the timber stand and the shrub layer scales, and because the number of subplots with coppice was different from the number of subplots where the information on the timber stand was available, two separate PCAs were performed, with each subplot weighted by its area. However, comparable results were obtained if only a single PCA on both data sets was performed. All analyses were performed using ADE (Ecological Data Analysis) software version 4 (Thioulouse et al. 1997). Spatial representation of subplots scores on PCA factorial axes (Goodall 1954) was performed using ARC VIEW 3.2 to first determine whether variation in vegetative composition was spatially structured, and then to establish the limits between such structures in the reserve. Jenks’ Goodness-of-Variance-Fit statistic (Jenks & Caspall 1971) was used to identify breakpoints between score classes. This iterative method minimises the sum of the variance within each class in order to detect the best groupings and patterns structuring the data. The number of classes was determined using score distribution. The frequency distribution of scores was plotted, and a non-parametric estimate of the probability density function was computed using S-PLUS software (Venables & Ripley 1997). The general trend of this function (number of peaks) allowed us to determine the number of PCA score classes to improve readability. To determine the proportion of total variability explained by our final classification, we calculated the ratio between the inter-class variability and total variability (for data sets). Many roads have been constructed within the reserve, which often constitute natural limits for roe deer home ranges (Hewison et al. 1998). Accordingly, we assessed major types of vegetation community using closest roads to delimit groups.

**Results**

**The timber stand scale**

Timber stands occurred in 574 out of the 629 subplots.
PCA on row profiles identified two axes. The first axis accounted for 65% of the total inertia and opposed oak and beech, the two main tree species in the reserve (256 oak-dominated and 224 beech-dominated subplots). The second axis of the PCA (accounting for 23% of the total inertia) opposed the oak-beech stands to pine (65 pine dominated out of the 574 subplots). Subplots distributed themselves in a triangular structure around the three main cover type categories (beech, oak and pine; Fig. 2). Most subplots were aggregated in the corners indicating that the gradient structures between timber stand types were weak.

Subplot scores plotted spatially relative to the first axis of the PCA (Fig. 3) supported the conclusion that the different timber stand types occurred in different areas. Beech occurred mostly in the south where 215 of 332 subplots (71% of the area) were beech dominated. Oak occurred in the north where oaks dominated 213 out of 242 subplots (90.7% of the area). Pine-dominated subplots were not aggregated and were distributed approximately equally in both parts of the reserve. Because there were relatively few pine-dominated subplots that did not show any spatial structure, we did not consider them when establishing our classification. The reserve was therefore divided into two groups (see Fig. 3). This classification accounted for 29.6% of the total variability of the timber stand data base.

The shrub layer
Coppices were present in only 237 of 621 subplots. PCA on row profiles performed on the subplots also identified a 2-axis structure. The first axis (accounting for 64% of the total inertia) opposed hornbeam-dominated to maple-dominated subplots. The second axis (accounting for 22% of the total inertia) opposed the latter subplots to beech dominated subplots. Subplots distributed themselves in a triangular structure around the three cover type categories (beech, hornbeam and maple; Fig. 4). Contrary to the timber stand results, most of the subplots were not aggregated in the corners indicating a gradient between the different types, particularly between hornbeam and maple dominated coppices.

The spatial representation of the subplot scores on the first axis of the PCA (Fig. 5) confirmed the spatial structure in the shrub layer: hornbeam-dominated coppice occurred essentially in the northeastern part while maple-dominated coppice occurred mostly in the northwestern part. Few subplots were dominated by beech (five subplots). We therefore did not take into account beech-dominated subplots when establishing our classification.

As subplots with coppices in the beech stand were not numerous (15 subplots) and differed in major cover type categories, we discarded them when dividing the reserve. Because the gradient structure was marked between coppice types, we used extreme score classes on the first
PCA axis to determine the limits of coppice types. Two types were thus distinguished (see Fig. 5): the maple-dominated coppice in the northwest and the hornbeam-dominated coppice in the northeastern part. This classification accounted for 32% of the total inertia in coppice data. Because of the gradient structure between coppice types, intra-group variability was relatively high: 64 of 73 subplots in the northwestern part (71% of the area) were dominated by maple, and 99 of 149 subplots in the northeastern part (67% of the area) were hornbeam-dominated subplots.

Discussion

Different and separate vegetation communities occurred in the Chizé reserve. Three major habitats can be identified from our analysis: the maple-dominated shrub layer in the northwestern part of the oak stand, the hornbeam-dominated shrub layer in the northeastern part of the oak stand, and the beech stand without shrub layer in the south. The high degree of spatial structure at the timber stand and the shrub layer scales could be expected because the forest has been managed by the ONF for more than 200 years. Our results fit previous conclusions based on data collected at the forest station scale on soil types and vegetation (Werno 1984). They also fit with results obtained on a data set collected in 2001 on the distribution of plant species between 0 and 1.20 m (Dray et al. 2002).

At the timber scale, our results validate those previously obtained on other vegetation data (Pettorelli et al. 2001) where two habitat types were recognised: the north zone corresponding approximately to the oak stand and the south zone corresponding to the beech stand. Roe deer access food items between 0 and 1.20 m (Duncan et al. 1998), so any difference in the dominant species at the stand scale was not expected to affect roe deer performance. However, the stand type was shown to affect the distribution of the preferred plant species by roe deer (Pettorelli et al. 2001, Dray et al. 2002). Moreover, this previous distinction was shown to affect roe deer performance, as we reported a significant 0.6 kg difference in winter body mass of fawns (Pettorelli et al. 2001), a significant 0.5 kg difference in adult female body mass and a significant 0.9 kg difference in adult male body mass (Pettorelli et al. 2002) between the heavier individuals living in the oak stand and those in the beech stand. We found no coppice in the beech stand. Coppices were identified as a major determinant of roe deer habitat quality, as they are directly linked to food availability and refuge possibilities (Cibien & Sempéré 1989, Mysterud et al. 1999). This may also explain why lower fawn and adult body masses were found in the beech stand (Pettorelli et al. 2001, 2002).
At the shrub layer scale, our results also underline the heterogeneity of the oak stand quality for roe deer. Hornbeam, which constitutes a principal (around 10% of summer diet by weight; Maizeret et al. 1991) and highly preferred (eaten 10 times as much as its availability in the reserve; Tixier et al. 1997) component of roe deer diet in spring and summer, is found primarily in the northeast part of the oak stand. Spring and summer are two periods of high-energy requirements for roe deer (Andersen et al. 1998). Females raise two fawns on average, and males defend their territory (Andersen et al. 1998). During this period, small cervids need to select food items that are highly digestible and rich in soluble carbohydrates (Maizeret & Tran Manh Sung 1984) like hornbeam (Tixier et al. 1997, Duncan et al. 1998). Considering the heterogeneity in expected habitat quality in the oak stand, the distinction of separate coppice types in the oak stand was expected to improve the importance of habitat quality for determining roe deer condition indexes like body mass. This was recently confirmed for fawn body mass, with a 1-kg difference (instead of 0.6 kg) between the heavier individuals living in the oak stand with hornbeam dominated coppices and those in the beech stand (Pettorelli et al. 2003).

Generally, approaches that concurrently use GIS and multivariate analyses are coupled with clustering methods in order to determine major vegetative communities. In our case, clustering methods were not conceivable for three different reasons. First, clustering methods generally do not consider spatial proximity. Spatially-constrained clustering procedures are available but were not usable because they generally apply to binomial or Poissonian data (generally provided by health science; Kulldorff 1997). Second, clustering methods (spatially constrained or not) do not consider natural frontiers like roads, which are major delimiters of roe deer home ranges (Hewison et al. 1998). Because our purpose was to define a major spatial structure of vegetation at a scale that is relevant for coupling vegetation and deer data, we needed to consider such components of deer habitats. Finally, there was a strong pattern in the vegetation community distribution and a simple spatial representation of PCA scores was clearly enough to define major coppices and timber stand types. The goal of our study was to define the major vegetation communities occurring in the reserve to better understand the role of habitat on deer population dynamics. Our aim was not to develop a robust method for identifying fine spatial vegetative structures.

This method could, however, suffer from the fact that delimitation of the different classes is not objective. This is particularly clear when dealing with gradient structures like the hornbeam-maple gradient occurring in our oak stand. But this simplification could also be viewed as the important point in our study. Deer populations have been continuously increasing in Europe over the last decades and some species like roe deer are difficult to count (Cederlund et al. 1998). Since 1990, researchers and managers have looked for index methods to replace counts as the basis for management decisions (Morellet et al. 2001). There is more and more evidence that habitat quality is a major determinant of body condition indexes and survival for deer species at both intrapopulation scales (Gaillard et al. 1993, Tilman & Kareiva 1997, Milner-Gulland et al. 2000, Pettorelli 2002). In some cases, not taking habitat substructure into account, even at the intra-population scale, could lead to a significant loss in revenue from management (e.g. by harvesting two contrasting subpopulations at the same rate; Milner-Gulland et al. 2000).

The need for coupling vegetation and animal data is growing, like the need for tools (Manly et al. 1993, Boyce & McDonald 1999). Considering the recent importance of habitat quality as a determinant of deer performance at the intra and inter-population scales, we believe that further research should be carried on the relationship between habitat type and structure and deer population dynamics, and that the coupling use of GIS and multivariate analysis could be helpful in that purpose. Forestry data in France are both abundant (descriptions of major plant cover per plot are available for nearly all managed forests) and sometimes underused. A first approach relating for example the forest potential estimated by occurrence of favourable vegetation communities to deer performance (e.g. body mass of hunted animals) could bring insights into how habitat composition affects deer performance at the national scale (Sperber 1975, Denis 1985).

We did not provide new methodology, as PCA, GIS and the combined use of both tools are known (Kadmon & Danin 1997, Ohmann & Spies 1998, Kadmon & Heller 1998, Guisan et al. 1999). What is new is the use of landscape ecology methodology for management purposes. The connection between vegetation and animal data is clearly a new challenge for ecologists and managers who need simple and efficient tools robust enough to integrate site particularities. We present a simple use of GIS and PCA, which provides simple and efficient results on the habitat structure at a relevant scale. We are aware that our approach could be improved (e.g. by developing spatially-constrained clustering methods suited for our type of data); such biological situations could constitute some very interesting departure points for biometry researchers.
Acknowledgements - we are grateful to the Office National des Forêts and to all field assistants and volunteers who spent time collecting data in the study site. Special thanks to Jean-Michel Gaillard, Daniel Chessel, Patrick Duncan, Stephane Villarubias, Vanessa Viera and people from the Office National des Forêts, Anne Loison, Atle Mysterud and Nils Chr. Stenseth, Erling Solberg, Noël and Nadine Guiñon. Thanks also to Nicolas Morellet and two other anonymous referees for valuable comments on previous versions of this manuscripts.

References


