EIGENANALYSIS OF SELECTION RATIOS FROM ANIMAL RADIO-TRACKING DATA

C. CALENGE1,2 AND A. B. DUFOUR1

1UMR CNRS 5558, Laboratoire de biom?trie, Universit? Claude Bernard Lyon 1, 69622 Villeurbanne Cedex, France

Abstract. The development of methods to analyze 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 test of Neu et al., can be used to determine whether habitat selection is significant. As well, 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 types of habitat are large, the biologist often has to deal with an excessively large number of measures. In this paper we present a new method, the eigenanalysis of selection ratios, that generalizes these two common methods within the framework of eigenanalyses. This method undertakes an additive linear partitioning of the White and Garrott statistic, so that the difference between habitat use and availability is maximized on the first factorial axes. The eigenanalysis of selection ratios is therefore optimal in habitat selection studies. Although we primarily consider the case where the habitat availability is the same for all animals (design II), we also extend this analysis 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.

Key words: chi-square; design II; design III; eigenanalysis; factorial analysis; habitat type; radio-tracking; selection ratios.

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 data 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 has been developed that 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 can 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. Studies with designs II and III 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: (1) whether selection is significant with regard to the habitat types considered, and (2) which habitat types are selected by animals.

Among the numerous methods available for question (1), 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 design-III studies. It consists of computing one chi-square value 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 (2) (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 J 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), can affect habitat selection and tend to increase its variability. That is, some habitats might be strongly selected by some animals and weakly selected by others. Therefore, averaging selection ratios should not be done without first investigating variability.

Eigenanalyses can be useful to achieve this objective (Greenacre 1984, Escoufier 1987). This family of methods groups all analyses that can be considered as extensions of principal component analysis (PCA). An eigenanalysis is characterized by a triplet of matrices: (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. The weighted PCA of the table assigns scores to its rows and its columns, so that the weighted sum of squares of the row scores, which we term the "inertia," 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, or correspondence analysis, 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 that select their habitat in a common way. We present an example application using the squirrel data set provided by Aebischer et al. (1993) and illustrate 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

\[
\begin{align*}
\sum_{j=1}^{J} u_{ij} &= \text{(total for habitat } i) \\
\sum_{i=1}^{I} u_{ij} &= \text{(total for animal } j) \\
\sum_{i=1}^{I} \sum_{j=1}^{J} u_{ij} &= \text{(total number of relocations of all animals)}
\end{align*}
\]

Let \( p_i \) be the proportion of available resource units in habitat type \( i \), also termed availability of habitat \( i \). 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:

\[
w_{ij} = \frac{u_{ij}}{u_{ij} / p_i}.
\]

Under the hypothesis that the animal \( j \) uses habitat type \( i \) randomly, this ratio is on average equal to 1. The larger this value, the stronger is the selection for habitat \( i \). Manly et al. (2002) proved that this ratio is the best estimate 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-square test proposed by White and Garrott (1990), which is carried out in two steps. First, the statistic

\[
(\text{chi-square})_j = \sum_{j=1}^{J} \frac{(u_{ij} - p_i u_{ij})^2}{p_i u_{ij}}
\]
is computed for each animal \( j \). Under the null hypothesis of random habitat use, this statistic is drawn from a chi-square distribution with \( I - 1 \) degrees of freedom.

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

\[
S = \sum_{j=1}^{J} \sum_{i=1}^{I} \frac{(u_{ij} - p_{i}u_{ij})^2}{p_{i}u_{ij}}.
\]

Note that this statistic may be written

\[
S = \sum_{j=1}^{J} \sum_{i=1}^{I} p_{i}u_{ij} \left( \frac{u_{ij} - p_{i}u_{ij}}{p_{i}u_{ij}} \right)^2
= \sum_{j=1}^{J} \sum_{i=1}^{I} p_{i}u_{ij} \left( \frac{u_{ij}}{p_{i}u_{ij}} - 1 \right)^2
\]

which is equivalent to

\[
S = \sum_{j=1}^{J} \sum_{i=1}^{I} p_{i}u_{ij}(w_{ij} - 1)^2. \tag{1}
\]

Under the random hypothesis of random habitat use by all animals, this statistic follows a chi-square distribution with \( J(I - 1) \) 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 \( u_{ij} \) values (Manly et al. 2002), i.e.,

\[
(S)_c = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(u_{ij} - u_{i+}u_{+j})^2}{u_{i+}u_{+j}}. \tag{2}
\]

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

Note that the classical correspondence analysis of the contingency table containing the \( u_{ij} \) values undertakes an additive partitioning of this statistic, therefore maximizing on the first axes the discrimination of animals according to their habitat use (Greenacre 1984, Thioulouse and Chessel 1992). This analysis assigns scores to animals and habitats, so that two animals with a similar profile of habitat use have a similar score and two animals with very different profiles of habitat use have very different scores. The amount of discrimination of animals on a given axis is measured by the corresponding eigenvalue of the analysis. The sum of these eigenvalues corresponds to the chi-square computed using Eq. 2.

### The Eigenanalysis of Selection Ratios

The classical correspondence analysis discriminates animals according to their habitat use. The aim of habitat selection studies is not to discriminate animals, but rather to emphasise the differences between their habitat use and habitat availability. Thus, instead of an eigenanalysis that undertakes an additive partitioning of the statistic computed in Eq. 2, which is a measure of the discrimination between animals, we want to develop an eigenanalysis partitioning the statistic computed in Eq. 1, which is a measure of habitat selection.

Let the \( I \times J \) matrix \( W \) (rows \( \times \) columns) contain the centered selection ratios:

\[
W = (w_{ij} - 1)_{i=1, \ldots, I, j=1, \ldots, J}.
\]

Each row of \( W \) contains the coordinates of an animal in the multidimensional space defined by habitat types (column space). Similarly, each column of \( W \) contains the coordinates of a habitat type in the multidimensional space defined by animals (row space). The centering of \( W \), i.e., the subtraction of 1 from the selection ratios, is essential for the analysis. Indeed, 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. We now search the vector in the column space for which orthogonal projections of animals are on average as far as possible from the origin. We also search the vector in the row space for which the orthogonal projections of habitat types are on average as far as possible from the origin. The solution can be found with the help of eigenanalyses (for the theory, see Escoufier [1987]; for examples, Thioulouse and Chessel [1992], Doledec and Chessel [1994], Doledec et al. [2000], Calenge et al. [2005]).

It can be shown that the solution of this problem is given by the eigenanalysis of the triplet \((W, P, D)\), where the \( J \times J \) matrix \( D = \text{Diag}(d_{1}, \ldots, d_{J}) \) is the weight matrix associated with the columns of \( W \), and the \( I \times I \) matrix \( P = \text{Diag}(p_{1}, \ldots, p_{I}) \) is the weight matrix associated with the rows of \( W \). This eigenanalysis corresponds to the principal component analysis of the \( W \) table weighted by rows and by columns (Escoufier 1987). This analysis 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 squared 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 greater importance is given to animals that relocate more. In fact, the use of these two weight matrices is necessary to achieve a very desirable property: it allows the linear and additive partitioning of the White and Garrott (1990) statistic of Eq. 1 by the analysis, a property that we demonstrate.

The total inertia of this analysis is equal to (Escoufier 1987) \( \text{Inertia} = \text{Tr}(W^t D W P^t) \), where \( W^t \) is the transpose of \( W \), and \( \text{Tr} \) is the trace operator, which returns the sum of the diagonal elements of the matrix.
on which it acts. This is equivalent to Inertia - $\sum_{i=1}^{t} p_i M_i^2 (w_i - 1)^2 = S$, where $S$ is the statistic computed in Eq. 1. Thus, the inertia of this analysis is equal to the White and Garrott (1990) statistic. Because the eigenanalyses allow the linear partitioning of the inertia, it is clear that this statistic is therefore maximized on the first axes. The first axes are those for which habitat selection is the highest. This analysis is optimal to study habitat selection with design II and can also be extended to design III in a straightforward way (see the Appendix).

**Application**

**Analysis of habitat selection by the squirrels**

We applied the eigenanalysis procedure 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-day period. Five habitat 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 that 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 (1990) chi-square test for habitat selection is highly significant ($S = 1563.9$, $df = 68$, $P < 0.001$), indicating a strong habitat selection by animals. Furthermore, the selection was not identical for all animals ($S_e = 321.7$, $df = 64$, $P < 0.001$). The eigenanalysis of selection ratios has been carried out to explain this heterogeneity.

The analysis produces two factors that predominately explain the White and Garrott 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: (1) some animals primarily used the larch habitat type, and used the mature and *Thuja* less frequently; and (2) others preferred the larch relative to the mature and young habitat types. 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.

**Ecological interpretation**

The variability in habitat selection highlighted by the analysis can 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 types, and the two patches are separated by 70 m. The home range size of the monitored squirrels, computed using the minimum convex polygon (Mohr 1947), covers on average 1.75 ha ($SE = 0.23$). A circle with the same area would have a radius of 75 m. As the home range is a life history trait that 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), although less intensively. Conversely, another group selected the young habitat and also used the neighbouring types.

The importance of patchiness of habitat types 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 can 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 eigenanalysis of selection ratios extended to design III (see the Appendix).

This analysis can be computed using the ade4 package for R (R Development Core Team 2005; freely available for download). The extension to design III can be carried out using the R package adehabitat (Calenge 2006).

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, includes within the framework of eigenanalyses the selection ratios and White and Garrott’s (1990) 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 eigenanalysis 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 avoidance by an animal of a 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, the stronger is the selection for the corresponding type.

The eigenanalysis 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 might even obscure important information on the individual variability in habitat selection, because this ordering can cover a wide range of situations. For example, an ordering of the habitat type “A greater than B” could 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 the present study. In the first explanation, the preferred habitat is
present in all home ranges, which could suggest that this habitat is necessary for the animals. Conversely, the second explanation shows that the animal can survive without the habitat A. This distinction can be of importance when setting a wildlife management or conservation policy.

The same problem might 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 modeling methods that test the differences between groups of animals (Aebischer et al. 1993, Manly et al. 2002). For example, Heisey (1985) showed the close connection that exists between the log-linear modeling and the selection ratios. Using such models, it is possible to test whether prespecified 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 of these analyses imply that the biologist knows a priori which variables potentially affect habitat selection. This 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 can 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 of the animals select the same habitat types, then use of the first axis of the analysis can explain most of the White and Garrott 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 their wide applicability. Because the eigenanalysis of selection ratios is essentially exploratory, it 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 by the number of animals. This is not to say that the number of individuals is not important, because it necessarily affects the inferential power of the technique. Another key quality of the eigenanalyses is that this methodology can be used to analyze a large diversity of designs (Escoufier 1987, Doledec et al. 2000, Dray et al. 2003, Thioulouse et al. 2004).

For all these reasons, this method family has been used for decades in other fields of ecology, in particular in plant ecology (Austin 1985, Ter Braak 1985, 1986, Thioulouse and Chessel 1992). Eigenanalyses, such as the correspondence analysis or the canonical correspondence analysis, have been widely used to discriminate plant species according to their distribution in sampling sites, therefore allowing the identification of the patterns in vegetation composition in numerous studies (Hill 1973, 1991, ter Braak and Prentice 1988, Austin 2002). Ölkand (1996) pointed out their great potential for the generation of hypotheses about vegetation–environment relationships. However, this family of methods is still scarcely used in animal ecology, which underlines the lack of communication between different ecological fields already noted by Austin (1999).

In the current analysis, we considered only one table of selection ratios (with animals in columns and habitat types in rows), but the analysis of selection ratios can also be developed for schemes of greater complexity. 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 among 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 eigenanalysis of selection ratios could 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.

Acknowledgments

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.

Literature Cited


EIGENANALYSIS OF SELECTION RATIOS 2355


APPENDIX

An extension of the eigenanalysis of selection ratios for design-III data (Ecological Archives E087-142-A1).