

ANALYZING FORAGING AND HABITAT USE THROUGH SELECTION FUNCTIONS

LYMAN L. McDONALD, BRYAN F. J. MANLY, AND CATHERINE M. RALEY

Abstract. Methods commonly used for study of natural selection in changing populations are useful in the evaluation of food and habitat selection. In particular, one can derive "selection functions" that allow estimation of the relative probability that a given food item (or habitat class) will be selected next, given equal access to the entire distribution of available food items (or habitat points). Procedures are available for the cases when food items are assigned to qualitative classes and when the items are characterized by measurement of quantitative variables on the item. The primary advantage of these analysis methods is that clear probabilistic statements can be made concerning the likelihood that each of several food or habitat types will be used. Illustrations are given by analysis of prey size selection by Wilson's Warblers (*Wilsonia pusilla*) and by Tree Swallows (*Tachycineta bicolor*).

Key Words: Food exploitation; fitness functions; natural selection; selection functions; selectivity indices; weighted distributions; habitat selection.

Evaluating food and habitat selection by a population of animals is an important aspect of ecological studies. Analysis methods in the literature often assume that food or habitat resources can be classified into one of several categories defined by the researcher (e.g., Carson and Peek 1987). The chi-square goodness-of-fit tests, the chi-square test of homogeneity with Bonferroni z-tests (Marcum and Loftsgaarden 1980), or a rank-order procedure involving the relative ranks of available and used food items (or habitat classes) (Johnson 1980, Alldredge and Ratti 1986), have been used to analyze availability and usage data. Other analysis procedures include common univariate (e.g., Quinney and Ankney 1985) and multivariate statistical methods. In particular, discriminant analysis has been used to make inferences toward food (habitat) selection in the multivariate case (Williams 1981).

When food items are classified into one of several categories, Manly's (1974) selectivity indices can be used to estimate relative probabilities of selection (Heisey 1985). In the study of natural selection on a population consisting of qualitatively distinct morphs, the selectivity indices are the relative "fitness" values of the morphs. In the present application, the indices are relative probabilities of selection of food items (or habitat classes) from categories defined by the researcher.

We consider the case where one or more quantitative variables $\{x_1, x_2, \dots, x_q\}$ can be measured on each available food item (or habitat point) and adopt parametric methods that have been developed for the study of natural selection in changing populations (Manly 1985). It is assumed that the relative probability of an animal selecting a food item (or a habitat point) given access to the entire population of available items (or habitat points) can be modeled by a function

of the variables, $w(x_1, x_2, \dots, x_q)$. This function is defined to be the selection function (it is called the fitness function in the study of natural selection). It is a function such that if $f(x_1, x_2, \dots, x_q)$ is the frequency of available items (points) with X-values $X_1 = x_1, X_2 = x_2, \dots, X_q = x_q$ before selection, then the expected frequency of these food items in the diet (habitat points used) is

$$w(x_1, x_2, \dots, x_q)f(x_1, x_2, \dots, x_q).$$

In other words, given access to all food in the population of available items, the probability that an individual item with X-values $X_1 = x_1, X_2 = x_2, \dots, X_q = x_q$ is selected is proportional to the selection function, $w(x_1, x_2, \dots, x_q)$. The value of the selection function can be thought of as a weighting factor that represents deviation of resource use from purely random use. The case when the function $w(x_1, x_2, \dots, x_q)$ is constant over the range of (x_1, x_2, \dots, x_q) corresponds to the situation in which selection is purely random. Selection functions are particularly applicable to the study of food (habitat) exploitation because of the ease of biological interpretation of relative probabilities.

One further requirement is that the population of resource items must be considered to be infinitely large or that samples of the available and used items are collected instantaneously. In practice, these requirements are never totally satisfied. They must be replaced by the assumption that the proportion of the population withdrawn by the sampling is so small that the basic characteristics of the original population remain unchanged. This requirement is not unique to the present method, and represents a major obstacle to the study of resource selection in natural systems.

The method is illustrated by its application to two data sets: (1) size selection of leafhoppers (family Cicadellidae) by Wilson's Warblers (*Wilsonia pusilla*) (Raley and Anderson, in press), and (2) prey selection by Tree Swallows (*Tachycineta bicolor*). Data for the second illustration are approximated from figures and tables in Quinney and Ankney (1985). Although not illustrated, application to the study of selection of habitat points in a study area is straightforward. For example, Harris (1986) studied nest site selection by Fernbirds (*Bowdleria punctata*) in Otago, New Zealand. He measured nine variables at each nest site and at the closest clump of vegetation to randomly selected points. If we assume the variables he measured influenced the probability of selection of clumps of vegetation for nest sites, and the Fernbirds had access to the entire distribution of vegetation clumps, then our method could be used to estimate the relative probability that a randomly located clump with $X_1 = x_1, X_2 = x_2, \dots, X_9 = x_9$ was selected as a nest site.

DEFINITIONS AND STATISTICAL MODELS

The distribution of $X = \{X_1, X_2, \dots, X_q\}$ for food items (habitat points) in a study area is defined to be the *distribution of available items* and is denoted $f(x)$. A subset of the items is used by a population of animals under study during a certain period of time. The distribution of X for the subset of items used is defined to be the *distribution of used items* and is denoted by $g(x)$. In the following, reference will be made only to selection of food items with the understanding that results are equally applicable to the study of habitat selection.

We follow the notation and models reviewed in Manly (1985:55-75), where applications to the theory of natural selection of animals in changing populations are considered. McDonald and Manly (1989) also used the mathematical and statistical results (of Manly [1985]) to develop a theory for calibration of biased sampling procedures.

We assume that animals are using food items from the available distribution such that the probability of selection of an item depends only on the variables measured in X and is proportional to the *selection function*,

$$w(x) = w(x_1, x_2, \dots, x_q). \quad (1)$$

The distribution of used items is proportional to the product of the selection function and the distribution of available items. When the proportionality constant is needed, it is denoted by

$$E_r(w[x]),$$

the expected value of the selection function with respect to the distribution of available items. The distribution of used items can then be written as

$$g(x) = w(x)f(x)/E_r(w[x]). \quad (2)$$

In general, the proportionality constant cannot be estimated unless the sizes of the populations of available items and used items are estimated. The distribution $g(x)$ is known as a "weighted distribution" in the mathematical statistics literature.

Assume that a random sample of n_0 items is selected from the study area and the vector of variables is measured on each to yield the sample of available items $\{x_{01}, x_{02}, \dots, x_{0n_0}\}$. Similarly, assume that a sample $\{x_{11}, x_{12}, \dots, x_{1n_1}\}$ of n_1 items is randomly obtained from the population of used food items.

Given the samples of available and used items, the selection function is estimated by appropriate formulae. One can then graph the relative probability of selection in two and three dimensions and test hypotheses concerning the significance of parameters in the selection function. If $w(x)$ does not depend on x (e.g., the constants in the following model are both zero), then the distributions of available and used items are equal. This case is equivalent to the conclusion that the animals are selecting food items from the population of available items at random. In the case that the relative probability of selection of points depends on the variables X_1, X_2, \dots, X_q , one can evaluate whether the selection results mainly in changes in the mean or the variance of a variable.

Manly (1985:55-75) reviewed cases where distributions of available and used items follow univariate normal, multivariate normal, or gamma distributions, and considered robust models for selection functions. He also developed a procedure for estimation of a general multivariate selection function without assuming any particular parametric form for the distributions. Here we give a brief outline of the use of the existing models by illustrating formulae for the case when both distributions are normal with a single variable in the selection function.

Assume that the variable X has a normal distribution with mean μ_0 and variance V_0 for the population of available items, and that the probability of selection of an item with $X = x$ is proportional to the selection function

$$w(x) = \exp\{kx + mx^2\}, \quad (3)$$

where k and m are constants. Under these assumptions, it is known that the distribution of used items will be normally distributed with mean μ_1 and variance V_1 , where

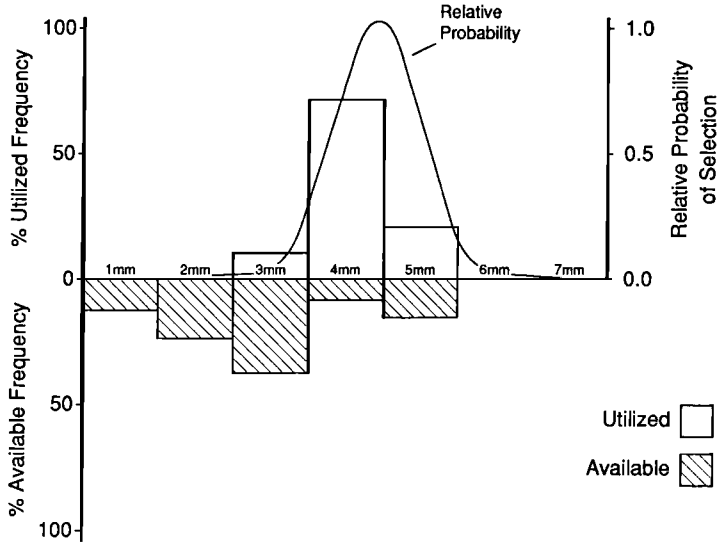


FIGURE 1. Relative probability of selection of leafhoppers by Wilson's Warblers plotted as a function of prey length. The graph is superimposed on the distributions of leafhopper lengths in the samples of available and used prey.

$$\mu_1 = (\mu_0 + kV_0)/(1 - 2mV_0), \text{ and} \quad (4)$$

$$V_1 = V_0/(1 - 2mV_0). \quad (5)$$

These equations can be solved for the constants in the selection function to yield

$$k = (\mu_1/V_1) - (\mu_0/V_0), \text{ and} \quad (6)$$

$$m = ([1/V_0] - [1/V_1])/2. \quad (7)$$

Also, if the distribution of X is normal in both populations, then the selection function must be of the form $w(x) = \exp\{kx + mx^2\}$.

One can denote the usual sample means and variances by $(\hat{\mu}_0, \hat{V}_0)$ and $(\hat{\mu}_1, \hat{V}_1)$ for the samples of available and used items, respectively. The reciprocal of the sample variance should be adjusted slightly when used to estimate the reciprocal of the corresponding parameter. For the two samples, $j = 0, 1$, let

$$\hat{B}_j = (n_j - 3)/(n_j \hat{V}_j), \quad (8)$$

with estimated variance

$$\text{var}(\hat{B}_j) = 2(\hat{B}_j)^2/(n_j - 5), \quad (9)$$

and let

$$\hat{a}_j = \hat{\mu}_j \hat{B}_j, \quad (10)$$

with estimated variance

$$\text{var}(\hat{a}_j) = (\hat{B}_j/n_j) + (\hat{\mu}_j)^2(\text{var}[\hat{B}_j]). \quad (11)$$

The estimators of k and m are

$$\hat{k} = (\hat{a}_1 - \hat{a}_0), \quad (12)$$

with estimated variance

$$\text{var}(\hat{k}) = \text{var}(\hat{a}_1) + \text{var}(\hat{a}_0) \quad (13)$$

and

$$\hat{m} = (\hat{B}_0 - \hat{B}_1)/2, \quad (14)$$

with estimated variance

$$\text{var}(\hat{m}) = (\text{var}[\hat{B}_0] + \text{var}[\hat{B}_1])/4. \quad (15)$$

ILLUSTRATIONS

We first consider a subset of data analyzed by Raley and Anderson (in press), who sought to quantify the relationship between availability and use of invertebrate food resources by riparian birds. They used Johnson's ranking procedure (Johnson 1980) to compare availability and selection of ten orders of invertebrates by the bird community. They evaluated size selection using the Kolmogorov-Smirnov two-sample test to compare the distributions of available and used items. We consider selection of one family of insects (leafhoppers) by one species of bird (Wilson's Warblers) during 15 June to 12 July 1986 (Fig. 1).

The distributions of available and used items are approximately symmetric and the assumption of normality of the lengths of leafhoppers in the populations will be made for this illustration. Under this assumption, and given that the birds have access to the entire distribution of

TABLE 1. FREQUENCY DISTRIBUTIONS OF PREY LENGTHS (MM) WITH SAMPLE MEANS AND STANDARD DEVIATIONS IN THE SAMPLES OF AVAILABLE AND USED PREY (APPROXIMATED FROM QUINNEY AND ANKNEY 1985)

Length (mm)	Available (frequency)	Used (frequency)
1	216	44
2	708	133
3	765	208
4	401	347
5	676	493
6	444	376
7	132	208
8	132	94
9	34	38
10	34	56
\bar{X}	4.01 mm	5.08 mm
SD	1.94 mm	1.89 mm

available leafhoppers, the estimated values for the constants in the selection function are

$$m = -1.4534 \text{ with } SE(m) = 0.02005, \text{ and} \\ k = 12.7481 \text{ with } SE(k) = 1.6565.$$

Both estimates are large with respect to their standard errors, indicating that both the mean and variance of used items are different from the mean and variance of available items. The estimated selection function (scaled by dividing by the largest selection value $\exp[27.9351]$) is

$$w(x) = \exp(12.7481x - 1.4534x^2 - 27.9351).$$

We divide by the constant $\exp(27.9351)$ so that the relative probability of selection of the most "preferred" length 4.5 mm is 1.0. Under the stated assumptions, the relative probability of selection was strongest for leafhoppers of length 4.5 mm by Wilson's Warblers during the period 15 June to 12 July 1986 (Fig. 1). In comparison, a leafhopper of length 3.5 mm was selected approximately one-third as often, whereas a leafhopper of length 5.5 mm was selected approximately one-fifth as often.

Quinney and Ankney (1985) reported size of prey selected (orders Diptera and Homoptera) by Tree Swallows. Their primary objective was to draw conclusions concerning optimal foraging theory. Data from one of their study sites, the sewage lagoon, were approximated from their Figure 1 and Table 3 (Quinney and Ankney 1985) for our second illustration (Table 1).

There was a significant shift in the mean length of used prey compared to available prey, but no significant ($P > 0.05$) difference in the variances. The pooled estimate of the common variance for the two distributions is $\hat{V} = 3.70$. Again, the distributions are approximately normal and es-

timination of the selection function follows the theory reported in eq. (3) to (15). From eq. (14), the constant m is judged to be zero because the variances are not significantly different. Using the common variance, $\hat{V} = 3.70$, in eq. (12) the constant k is estimated to be

$$k = 0.2892 \text{ with } SE(k) = 0.03469.$$

The relative probability of selection of a prey item of length x given access to the entire distribution of available insects is estimated by the selection function

$$w(x) = \exp(0.2892x - 2.892).$$

Again, the original function $w(x) = (0.2892)x$ has been scaled by dividing by $\exp(2.892)$ so that the relative probability of selection of the most "preferred" length 10 mm is the number 1.0 (Fig. 2). An insect of length 9 mm is selected with approximately three-fourths the probability of selection of an insect of length 10 mm, while an insect of length 5 mm has approximately one-third the chance of being eaten.

We consider the classes in Table 1 as qualitative to illustrate the computation of Manly's selectivity indices (Manly 1974) for qualitative variables. The selectivity indices are relative probabilities of selection and are interpreted in exactly the same manner as a particular value of the quantitative selection function considered above. Estimators of the selectivity indices (relative probabilities of selection) for length $x = 1, 2, 3, \dots, 10$ mm, are

$$w(x) = f_{ux}/f_{ax}, \quad (16)$$

where f_{ux} is the frequency of length x insects in the sample of used prey, and f_{ax} is the frequency of length x insects in the sample of available prey. If the frequency of length x insects in the sample of available prey is $f_{ax} = 0$, then some of the classes must be combined to avoid division by zero. Heisey (1985) recently developed procedures to estimate the selectivity indices under various hypotheses concerning how they depend on other attributes such as age and sex of the birds. Table 2 contains the selectivity indices for the 10 lengths of prey computed by eq. (16) and standardized by multiplying by $0.6071 = 1.0/w(10)$, so that the relative probability of selection of an insect of length 10 mm is 1.0.

The selectivity indices in Table 2 are presented only for the purpose of illustrating the analysis of qualitative classifications for habitat points or food items. The approximations of the frequencies of length 7-10 mm insects from Figure 1 and Table 3 of Quinney and Ankney (1985) are not very precise. Errors in the approximations will influence the selectivity indices in Table 2

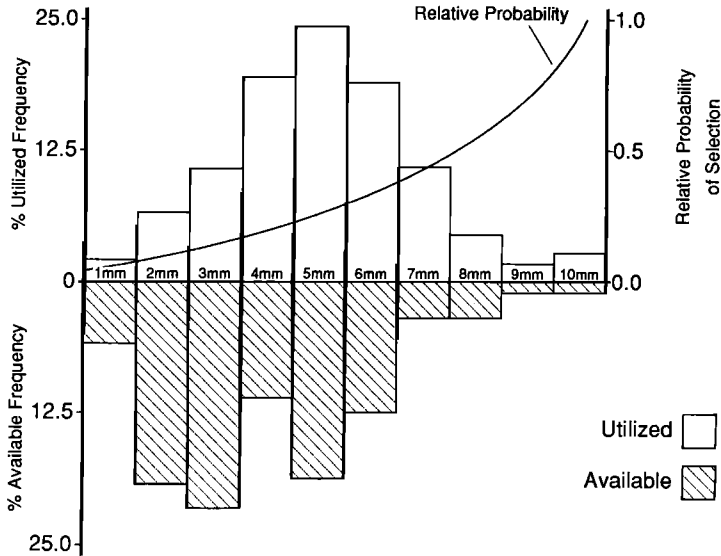


FIGURE 2. Relative probability of selection of insects by Tree Swallows plotted as a function of prey length. The graph is superimposed on the distributions of insect lengths in the samples of available and used prey.

relatively more than the selection function values graphed in Figure 2.

Our analyses agree with the conclusions of Quinney and Ankney (1985): "Swallows were selective in the sizes of insects that they captured. . . . distribution of sizes of insects captured by the birds from the 10 size classes was significantly different from the distribution of sizes present The two smallest classes (1 and 2 mm) were most underrepresented in the diets . . . in relation to their abundance in the nets."

DISCUSSION

If items are assigned to classes (perhaps purely qualitative), chi-square analysis sometimes leads to interpretation problems (Johnson 1980). If there are relatively large classes of available items that are seldom used, then the relationship of the expected frequency of use to observed frequency of use depends on whether or not the researcher includes those classes in the analysis. One of Johnson's objectives was to develop a procedure to overcome this problem. A major advantage of the selectivity indices is that the estimates are not sensitive to whether or not the large available but seldom used classes are included. For example, the 1 mm insects in Table 1 could be dropped and the selectivity indices (relative to selection of 10 mm insects) in Table 2 do not change. Another common practice is to standardize the selectivity indices so that they sum to 1.0, in which case dropping the large but seldom used class will have little effect.

The power of the selectivity indices and the selection function (univariate and multivariate) is that they provide clear probabilistic statements in the study of resource use. There may be several confounding reasons why an estimated selection function assigns significantly different relative probabilities of selection to items with different characteristics. Discussion of the biological reasons is beyond the scope of this paper; however, violation of the required assumptions may also contribute to a seemingly significant result or hide an important result. The basic assumptions required are: (1) the correct models have been used for the selection function and the distributions of available and utilized samples, (2) sampling

TABLE 2. SELECTIVITY INDICES COMPUTED AS IF THE CLASSES IN TABLE 1 WERE QUALITATIVE. INDICES ARE STANDARDIZED SO THAT THE INDEX FOR LENGTH 10 MM IS 1.000

Length (mm)	Selectivity index
1	0.124
2	0.114
3	0.165
4	0.525
5	0.443
6	0.514
7	0.957
8	0.432
9	0.679
10	1.000

is instantaneous, so that the characteristics of the distributions do not change during the sampling period, (3) the basic sampling unit is the individual food item (habitat point) and must be independently collected, and (4) the researcher has identified those classes or measured those variables that actually influence the probability of selection. An alternative to assumption (3) is that estimates of the selection function are independently replicated over the study area (or time).

The assumption that the correct models have been used can be tested with one of the common statistical tests (e.g., chi-square goodness-of-fit); however, power will be low unless the sample sizes are large. In the illustrations presented the distributions were approximately symmetric except for the available distribution of prey lengths to Tree Swallows (a chi-square goodness-of-fit test comes close to rejecting the hypothesis of normality for this data set). Although the theory is not yet available to defend a formal statement, we think that mild departures from normality will have little effect on the estimated selection function presented in eq. (3).

Estimation procedures are available for one more parametric situation in addition to normality. This case arises when both the distributions are skewed in the same direction and gamma distributions satisfy assumption (1). Estimation formulae appear in Manly (1977).

O'Donald's general quadratic selection function (reported in Manly [1985]) can be fitted between any univariate distributions, normal or otherwise. Use of this selection function diminishes the importance of assumption (1). It is also simple to use, but has disadvantages because the quadratic function may not fit the selection function over its entire range (in fact, it may be negative for extreme values of the variables) and procedures for drawing statistical inferences are not available. A second procedure that does not make assumptions about the parametric form of the distributions is Manly's general multivariate technique (Manly 1985). A robust exponential model is fitted to the selection function. The estimation of fitness functions in the study of natural selection assumes that two or more samples are available over time from the changing population. In the estimation of selection functions for food or habitat exploitation only two samples are available. Further theoretical research is needed to evaluate the statistical properties of Manly's general multivariate technique for the case of two samples and to develop new numerical (nonparametric) estimation procedures.

The second assumption (that sampling is instantaneous) is difficult to satisfy in practice. Whether or not this assumption is reasonable

also depends on the "basic sampling unit," discussed below. If the population of available units is very large with respect to the population of used items during the sampling period, then this assumption is not critical. Inferences will be to the "average" distributions during the study. But, if the population of available units (points) is limited and "preferred" units are quickly selected, utilization is changing the available distribution and the present techniques are not applicable. Again, further theoretical research is needed to evaluate selection from a changing population.

In the illustration of selection of leafhoppers by Wilson's Warblers, the population of leafhoppers was judged to be very large with respect to use by warblers. However, if this were not the case and the shape of the available distribution was changing, the low relative frequency of 4 mm leafhoppers (Fig. 1) might be exaggerated by selection for insects of that length. Consequently, this would exaggerate the estimated height of the selection function in this region.

Our analyses were made under the assumption that the samples of food are equivalent to random samples from the populations (i.e., food items in the samples are independently collected). It is rare that studies can be designed so that individual food items are the basic sampling units. It is important to keep track of the different sources of sampling variance and to avoid the infamous pseudo-replication problems (Hurlbert 1984). This is exceptionally difficult in studies of food selection because different collection methods are generally used to obtain the samples of available and used items. One procedure followed by Raley and Anderson (in press) was to also collect invertebrates from the shrub on which the bird was observed to feed. If birds are independently located, then it may be appropriate to consider the bird as the basic sampling unit. The sample of available invertebrates from a unique shrub is paired with the sample of used invertebrates taken from the bird. If there are sufficient numbers of insects in the paired samples, then replicate estimates of the parameters in the selection function could be obtained for each bird. Alternatively, estimates of the parameters in the selection function might be replicated over some larger unit of space or time. For example, data for birds collected in one of several independently sampled quadrats might be pooled. The selection function could be fitted in each quadrat to obtain replicate estimates of the parameters. Statistical inferences toward the mean values for the parameters of the "average" selection function over the study area (or time) would be made by considering the variances of the replicate estimates.

In the study of habitat selection we envision that the basic sampling unit will be a point or a small area associated with a point. A random sample of available points (or a systematic sample with random starts) might be selected with a vector of variables to be measured at each point. Points in this sample that happen to be used are left in the sample of available points. Obtaining a random sample of used points will be more difficult and highly dependent on the particular application.

ACKNOWLEDGMENTS

Work on this project was partially supported by a grant from the Beverly Fund, University of Otago, Dunedin, New Zealand. Field studies were partially supported by the U.S.D.A. Forest Service, Rocky Mountain Forest and Range Experiment Station, Laramie, Wyoming. Significant improvements in the manuscript are due to the review by the referees: Douglas H. Johnson, Martin G. Raphael, Dana L. Thomas, and the editors. Their contributions are gratefully acknowledged.

