

SOME COMMENTS ON NICHE ANALYSIS IN CANONICAL SPACE¹

BRUCE A. CARNES

Argonne National Laboratory, Argonne, Illinois 60439 USA.

AND

NORMAN A. SLADE

Museum of Natural History and Department of Systematics and Ecology,
University of Kansas, Lawrence, Kansas 66045 USA.

Abstract. Discriminant functions have been used to identify axes of niche separation, and dispersion of locations on these axes have, in turn, been interpreted as representing niche width and specialization. We discuss the influence of sampling scheme and choice of niche measures on the results of an analysis conducted in canonical space. Specifically, habitat variables measured at random locations should provide a more representative measure of habitat available than would recording habitat only where animals are observed. Species presence information alone tends to bias the mean habitat toward the one where the most common species is found. If niche specialization is to be measured by distance from the overall mean habitat, each sampling station should receive equal weighting. The most straightforward measure of habitat breadth seems to be the variance or standard deviation of canonical scores. The mean squared distance from the species centroid can then be used as a comparable multidimensional breadth measure. Finally, we suggest that niche overlap be measured via a joint probability density function rather than by area of overlap of concentration ellipses.

Key words: canonical space; niche breadth; niche overlap; niche position; species absence.

INTRODUCTION

Hutchinson's (1957) concept of the ecological niche as a n -dimensional hypervolume invites the use of multivariate statistical techniques. Such methods allow many, sometimes highly correlated, variables to be displayed along independent axes representing what Rotenberry and Wiens (1980) call "proximate" niche dimensions. Both principal-component (PCA) and multiple discriminant-function analysis (DFA) have been used in niche analysis, with the latter technique emphasizing species differences (see discussion following Williams [1981]) and minimizing niche overlap. Rotenberry and Wiens (1980, 1981) raise some important philosophical questions underlying the choice of technique. Would one choose DFA under the hypothesis of independent assemblages of species (Gleasonian communities) as readily as under the hypothesis of competitively structured communities (J. Wiens, *personal communication*)? These issues should be addressed in the initial stage of any niche analysis.

The use of DFA seems to follow from Levins (1968: 48), "Thus niche dimensionality refers not to the number of biologically relevant factors in the environment, . . . but to the number of factors which serve to separate species." Cody (1968) and James (1971) used DFA to separate bird species according to habitat characteristics, but recent applications to niche analysis appear to be based primarily on the insightful pa-

pers of Green (1971, 1974) and Shugart and Patten (1972). Like Cody, Green (1971) used DFA to identify significant ecological factors separating species in multidimensional space, but he emphasized the niche concept more strongly. Green (1974) suggested the standard deviation of discriminant scores on an axis as a measure of niche breadth, and the percent overlap of 50% probability ellipses as a multivariate measure of niche overlap. Following Green's method, M'Closkey (1976) also measured niche breadth as the within-species standard deviation of discriminant scores along a single axis. Shugart and Patten (1972) developed a different series of niche measures based on generalized distances and discriminant functions. Dueser and Shugart (1978, 1979) extended this work and suggested average distance of the observations for a species (\bar{d}_i) from the origin of discriminant space as a measure of "niche position or exploitation specialty relative to the average microhabitat sampled" and the coefficient of variation (S_{d_i}/\bar{d}_i) of the distances from the discriminant origin as a measure of niche breadth.

In this paper we discuss the use of DFA as developed by these authors. We address the problems of selection of habitat vectors for analysis, estimation of niche or habitat specialization, niche width, and niche overlap. We have focused on DFA, not because of personal preference over PCA, but because DFA has been more frequently used in niche analysis to date. However, our discussion of niche metrics applies equally well to PCA, and we will use the term canonical space and axes when referring to results applicable to either PCA or DFA.

¹ Manuscript received 5 October 1981; revised 2 April 1982; accepted 5 April 1982.

Selection of habitat vectors

One consideration in multivariate analysis of habitat use or niche space is whether to sample environmental variables at sites determined randomly (or in a grid) or to record environmental variables only when one observes an organism of interest. Rotenberry and Wiens (1980, 1981) used the former method for generating a set of habitat vectors to be analyzed by PCA, but they also compared the habitat axes generated by random sampling with the results of PCA performed on vectors representing species means (i.e., habitat recorded only where species were present with each species weighted equally) and found close correspondence. Cody (1968, 1978), James (1971), Green (1971, 1974), Shugart and Patten (1972), M'Closkey (1976), and Dueser and Shugart (1978, 1979) all used the latter method (i.e., centers or singing perches within territories for birds and successful collecting sites for molluscs and mammals) to identify locations for habitat sampling. Depending on whether one analyzes the original observations or species means, this sampling approach produces a set of habitat vectors weighted by frequency of occurrence of individuals or species, and could lead to a biased estimate of available habitat. To justify collecting environmental data only where a mollusc was found, Green (1971) cited Hutchinson's (1968) reasoning that presence of a species in a sample was more informative than absence. Presence indicated a species could live at a site, but absence was ambiguous; absence might have arisen because the site was unsuitable, because the species had not yet colonized an otherwise suitable area, or because the species actually was present but missed in the sample. The distinction between interpretability of presence and absence may be meaningful for sedentary forms such as molluscs or when applied on a macrogeographic scale, but it appears less meaningful when applied to microgeographic habitat selection of vertebrates. Mobile animals may be found in marginal or completely unsuitable habitat, especially during seasons of high dispersal; thus, presence is not always indicative of habitat suitability (Carey 1981, Wiens and Rotenberry 1981). In effect, the species may be included or excluded in the habitat sample by chance. Even sedentary forms may settle and survive in areas where they cannot reproduce, thus existing in habitat outside their niche, *sensu* Maguire (1973), until they die.

Ignoring species absence and perhaps biasing the estimated mean habitat available can affect the interpretation of a niche analysis using DFA (or PCA). The origin of canonical space represents the grand mean of discriminant (or factor) scores. When species vary in relative abundance and habitat vectors are recorded only when individuals are observed, common species contribute more to the determination of the grand mean than do rare species and, therefore, are more

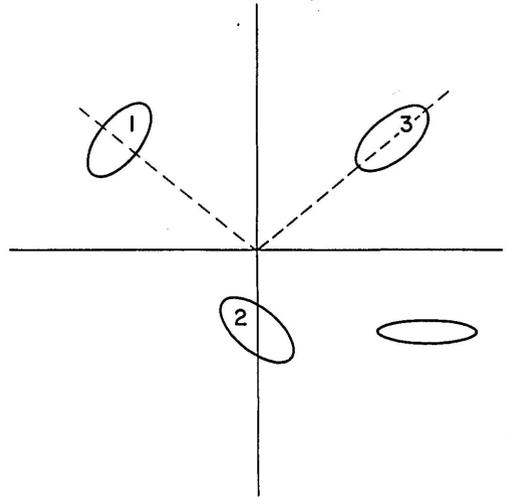


FIG. 1. The effect of differences in sample size on locations of concentration ellipses of three hypothetical species in canonical space determined by recording habitat where individuals were observed. Species 2 was twice as abundant as species 1 and 3, and showed habitat covariances of opposite sign. Species are numbered in rank order of variance in distance to the origin, S^2_{di} , 1 having the smallest variance. Mean squared distances from group centroids and the determinants of variance-covariance matrices (generalized variances) are equal for all three species. The unnumbered ellipse represents a smaller generalized variance but larger mean squared distance than the numbered ellipses.

likely to be located near the origin of canonical space. The increased probability of proximity to the origin will have a direct impact on the measures of niche position (mean distance to the origin, \bar{d}_i) and niche breadth (coefficient of variation of distances, S_{d_i}/\bar{d}_i) proposed by Dueser and Shugart (1979), who considered the discriminant origin to be the average microhabitat available. Common species will tend to be seen as broad-niched generalists, and rare species will be narrow-niched specialists, even if all species have equally variant habitat (or resource) utilization functions (Fig. 1).

The recording of habitat variables at each sampling site for each observation period (e.g., each trap-night in mammal studies) weights each sampling site equally. The origin of canonical space is then the mean of all sites sampled, rather than a mean of vectors weighted by frequency of occurrence of animal species. The inclusion of species absence should eliminate some of the effects of sample size on the niche measures of Dueser and Shugart (1979) because the origin of canonical space better estimates the average microhabitat available. An additional advantage is that some sampling sites will not be used, some will be used only occasionally, and others frequently; intensity of use can then be incorporated into describing

TABLE 1. Summary of the rank order distance from the discriminant origin (2) in the 360 simulated runs with unequal (5) and equal (6) sample sizes. Column 1 is the number of switches in rank order necessary to achieve the order (1234). Column 2 represents the 24 possible rank orders. Columns 3 and 4 represent the proportion (3) and number (4) expected under the equally likely hypothesis.

(1) No. of switches	(2) Order of distance	(3) Proportion expected	(4) Number expected	(5) Number (\neq)	(6) observed (=)
0	1234	1/24	15	37	11
1	1243	3/24	45	81	48
	1324				
	2134				
2	1342	5/24	75	98	71
	1423				
	2143				
	2314				
	3124				
3	1432	6/24	90	69	81
	2341				
	2413				
	3142				
	3214				
	4123				
4	2431	5/24	75	51	96
	3241				
	3412				
	4132				
	4213				
5	3421	3/24	45	22	43
	4231				
	4312				
6	4321	1/24	15	2	10
		1	360	360	360

niche boundaries. The disadvantage is that it becomes more difficult to separate used and unused habitat.

MONTE CARLO SIMULATIONS

To demonstrate the effect of sample size on discriminant analysis of samples from identical populations, we conducted a series of Monte Carlo simulations. We used four groups (species) and two variables (habitat or resource dimensions), but the procedure and results can be generalized to cases of higher dimension. Unequal sample sizes of $n_i = 40, 90, 140,$ and 190 were selected to represent the rarest through commonest species. A minimum sample size of 40 was selected to meet the recommended lower limit for discriminant analysis (Tatsuoka 1970). To simulate equal resource utilization functions and to meet the assumptions of DFA, we used IMSL (1979) routines to generate n_i random bivariate normal variates with mean $\mu = (0,0)$ and covariance matrix $\Sigma = I$ (identity matrix) as the vectors of habitat utilization for each species. Random uniform (0, 10) variates, representing the mean of the i^{th} species on the j^{th} habitat axis, were added to each element of the respective individual vectors, thereby

generating a set of bivariate normal vectors with means (μ_{i1}, μ_{i2}) and covariance matrix $\Sigma_i = I$ for each species ($i = 1,2,3,4$). This procedure was repeated 360 times with each simulation subjected to DFA (IMSL 1979). A second set of 360 simulations was generated with equal sample sizes ($n_i = 115$) for all four species. Thus 720 four-group DFA were performed, one for each set of random samples. Niche positions were calculated in discriminant two-space for each set.

Within the analysis of each simulation, the \bar{d}_i were ordered with the group farthest from the origin of discriminant space in the first position. The frequencies of the 24 ($=4!$) distinct arrangements were tabulated for equal and unequal sample sizes (Table 1). Under the null hypothesis that sample size had no effect, each rank order of distance from the origin was equally likely. Under the alternate hypothesis that rare species should have larger \bar{d}_i , the sequence (1234) should have occurred more frequently than others. We first tested whether the sequence (1234) occurred more frequently than expected under the null hypothesis. The test statistic is binomial, with $n = 360$ and $p = 1/24$. The alternate hypothesis implies $p > 1/24$ and a one-tailed rejection region, $x \geq 22$, of size $\alpha = .049$, where x is the frequency of the event (1234) in 360 trials. The 37 occurrences of the event (1234) in the simulations involving unequal sample sizes provided strong evidence for rejection ($P < .0001$) of the null hypothesis. For the equal sample size simulations, an observed value of $x = 11$ was not sufficient evidence to reject the null hypothesis of $p = 1/24$.

Next, the outcomes of the simulations were tabulated according to the number of switches (s) in rank order of \bar{d}_i necessary to achieve the order (1234). That is, if the observed order was (1234), $s = 0$; for observed orders of (1243), (1324), or (2134), $s = 1$, etc. The cumulative relative frequencies of observed and expected values (Table 1) demonstrate that rank orders close to (1234) occurred much more frequently than expected for the unequal sample size simulations but not for those with equal sample sizes. The Kolmogorov goodness-of-fit test statistic d_{\max} was highly significant for unequal n_i ($P \approx .001$) but insignificant for equal sample sizes ($P \approx .65$).

Thus, there is clear evidence that sample size does affect group position in sample canonical space (a similar result could be obtained for PCA), with rare species tending to be farther from the origin when determined only by species presence. If one accepts Dueser and Shugart's (1979) reasoning that \bar{d}_i is a measure of niche specialization, one should attempt to determine a less-biased estimate of the mean habitat available. Use of species means or subsamples from more common species will eliminate the sample size bias, but presence data alone may still give a biased estimate of mean habitat available. One should use a random or stratified random-sampling scheme to ob-

tain a representative estimate of mean habitat available.

Niche breadth along a single dimension

Since niche position or exploitation speciality is affected by sample size, it follows that niche breadth as measured by the coefficient of variation is also affected by sample size. The coefficient of variation (S_a/\bar{d}_i) is an inverse function of \bar{d}_i ; therefore, rare species are more likely to have narrow estimated niche breadths using this criterion because \bar{d}_i will tend to be large relative to the other species in the analysis even with equally variant utilization functions. In a subset of 50 simulations with unequal n_i , the rank order of coefficients of variation was the reverse of rank-ordered mean distances 33 times and differed by only one switch 14 times.

Dueser and Shugart (1979) selected the coefficient of variation because *Blarina brevicauda* and *Ochrotomys nuttalli* appeared to be microhabitat specialists relative to *Peromyscus leucopus*, yet they occupied larger areas in discriminant space. This was inconsistent with their impressions from fieldwork and with the common assumption (e.g., Sugihara 1980) that relative abundance is determined by the proportion of resources (niche space) used by a species. The apparent paradox of a ubiquitous but narrow-niched species such as *Peromyscus* in this example is resolved by recognizing that variation in canonical space represents the variety of habitats occupied, not the abundance of a particular habitat or its spatial distribution in the environment. In his study of European Warblers, Cody (1978) presented data showing that even though *Silvia borin* had probability ellipses approximately five times the area of those of *Silvia curruca* in discriminant space, *S. borin* could potentially occupy less than twice as much of the environment as *S. curruca*. In even more-extreme cases, a species might be restricted to a single microhabitat, that is, a dot in discriminant space, but if that microhabitat were common and widely dispersed, so might be the species. McNaughton and Wolf (1970) illustrate this point, citing barnacles and cattails as examples. If abundance of resources were inversely related to the distance from the origin of canonical space as Shugart and Patten (1972) suggested, one might predict from optimal-foraging theory (Stenseth 1981) that those species that use average (and hence more abundant) resources would restrict their breadth of resource utilization.

Green (1974) considered niche breadth as the variety of habitats occupied, measured by the variation in discriminant scores. As such, variance or standard deviation of scores seem more straightforward indices of niche breadth and are preferable to the coefficient of variation of distances on two counts: (1) they have well-known statistical properties, and (2) they are less

influenced by differences in sample size between species.

Multidimensional niche breadth

As Dueser and Shugart (1979) recognized, difficulties arise in extending niche breadth to two or more dimensions. These are similar to some problems Cody (1974) addressed for niche overlap measures (e.g., whether multidimensional breadth should be sums or products of unidimensional measures). One could measure niche breadth multidimensionally as a function of the determinant of the covariance matrix (generalized variance; Morrison 1976) of species scores on two or more discriminant axes. Levins (1968) referred to the area of a planar cross section of niche space as niche breadth, and Green (1971) suggested the area of a species concentration ellipse or volume of the ellipsoid as a niche breadth measure. The area or volume of an ellipsoid is proportional to the square root of the generalized variance (Green and Carroll 1976). Area of an ellipse is a product measure of breadth and was used by Dueser and Shugart (1979) in calculating overlap. As Cody observed with respect to α 's, products are most representative when the dimensions are independent (i.e., all covariance terms equal zero), so that the generalized variance is equal to the product of individual variances. If the assumption made in DFA of homogeneity of within-species covariance matrices is valid, then all species will be represented by congruent concentration circles, spheres or spheroids in discriminant space. However, equality of within-species covariance matrices is unlikely (Green 1974, Dueser and Shugart 1979), and generalized variances can conceal variation in one or more dimensions for those species that are represented by elongate concentration ellipsoids (i.e., highly correlated scores on discriminant axes or large differences in standard deviations of scores [Fig. 1]).

Shugart and Patten (1972) used the rate constant for decrease in population density (or intensity of habitat use) with increased distance from the observed mean habitat for a species as an index, r , to niche width (the larger $|r|$, the narrower the niche). This is an appealing idea for two reasons. Their argument justifying r as a niche width measure did not involve any restrictive assumptions about the resource or habitat utilization functions, so it is quite general, and distances have the advantage of being applicable in multidimensional as well as unidimensional space. Nevertheless, it is unclear how distance should be measured and what function should be used. As discussed earlier, Dueser and Shugart (1979) used S_a/\bar{d}_i as a breadth measure, but they also mentioned $S^2_{a_i}$ as a niche width measure. In multidimensional space, $S^2_{a_i}$ (where d_i is measured from the origin) is sensitive to variation only in lengths of radii connecting observations to the origin (Fig. 1). If the scatter of points for a species does not

surround the origin, then $S^2_{d_i}$ (or S_{d_i}) varies with the orientation of the concentration ellipse. Congruent ellipses will produce different values of $S^2_{d_i}$; specifically, $S^2_{d_i}$ will be smallest when the major axis of the ellipse is perpendicular to a radius from the origin and will increase as the major axis rotates toward a radius. If the observations scatter around the origin, $S^2_{d_i}$ will be less affected by orientation but will also be influenced by the distance between the group centroid and the origin. These effects can be avoided by measuring distances from the species centroid rather than the origin, and by doing so one comes closer to the concept of Shugart and Patten (1972).

J. Rakestraw (*personal communication*) has suggested the mean distance in discriminant space from observations to the species centroid as a multidimensional measure of niche breadth. This measure has the advantages of being relatively unaffected by the position of the origin or by sample size and of being comparable to the measures of Green (1974) and M'Closkey (1976). However, we prefer using the mean squared distance, $\sum d^2_{ij}/n_i$, which can be shown (see Appendix) to equal the sum of the maximum-likelihood estimators (n_i rather than $[n_i - 1]$ in the denominator) of variances of scores on the individual discriminant axes. If individual scores are standardized multivariate normal variates, then squared distances from the species centroid have the advantage of being approximate chi-square variates, so statistical comparisons of niche width could be made using F ratios. Both mean distance and mean squared distance reflect the sum of variation within species on the discriminant axes and are analogous to Cody's (1974) summation α .

Multidimensional niche overlap

Green (1971) suggested that niche overlap could be calculated as the overlap of probability (resource utilization) distributions in discriminant space but noted that visual inspection of overlap of 50% probability ellipses would usually be adequate. Dueser and Shugart (1979) followed this suggestion and measured the planar area of overlap of concentration ellipses relative to their total area. Concentration ellipses are generated from an estimate of the exponent of the multivariate normal density function for discriminant scores $(x - \bar{x}_i)'s_i^{-1}(x - \bar{x}_i)$, where s_i is the estimated covariance matrix, and \bar{x}_i is the estimated mean of discriminant scores for the i^{th} species. This quadratic form specifies the equation of an ellipsoid in p -dimensional space when set equal to a positive constant (Morrison 1967). Thus, the creation of concentration ellipses or ellipsoids rests on approximately multivariate normal utilization functions; whereas, the overlap measure proposed by Dueser and Shugart (1979) implies uniform utilization density within the ellipsoid. Departures from uniform density within the ellipsoid could result in misleading estimates of niche overlap. Harner

and Whitmore (1977) give simple formulae for two niche overlap measures in discriminant space when the assumption of homogeneity of within-species covariance matrices is valid. It is unfortunate that such instances may be rare with ecological data (Green 1974, Dueser and Shugart 1979). Recently, Maurer (1982) has generalized the results of Harner and Whitmore (1972) and extended them to the results of PCA.

In conclusion, Green (1974) and Dueser and Shugart (1978, 1979) have presented ecologists with a potentially valuable set of analytic tools. The specific measures of niche specialization and niche width used by Dueser and Shugart (1979) are affected by variation in number of samples per species, but these effects can be avoided either by random or stratified random sampling of habitat characteristics, regardless of species presence. The use of variance or standard deviation of canonical scores seems preferable to the use of the coefficient of variation as a niche breadth index and can easily be extended to multidimensional breadth by using the determinant of the covariance matrix or the mean of distances or squared distances from the group centroid. Finally, planar overlap of concentration ellipses may provide an index to niche overlap, but more representative measures should not be too difficult to calculate, and would be in closer accord with the distributional assumptions of discriminant analysis. Such uses of multivariate methods will lead to advances in understanding community structure and organization in complex environments.

ACKNOWLEDGMENTS

This work was partially supported by University of Kansas General Research Grant 3534-0038. We wish to thank Howard Levenson, Gregory Glass, James Rakestraw, John Wiens, and two anonymous reviewers for their helpful comments on earlier versions of this manuscript. Thanks are also extended to Michael E. Ginevan for his valuable input on the Monte Carlo simulations.

LITERATURE CITED

- Cary, A. B. 1981. Multivariate analysis of niche, habitat and ecotopes. Pages 104-113 in D. C. Capen, editor. The use of multivariate statistics in studies of wildlife habitat. United States Forest Service General Technical Report RM-87, Rocky Mountain Forest and Range Experiment Station, Fort Collins, Colorado, USA.
- Cody, M. L. 1968. On the methods of resource division in grassland bird communities. *American Naturalist* 102:107-147.
- . 1974. Competition and the structure of bird communities. Princeton University Press, Princeton, New Jersey, USA.
- . 1978. Habitat selection and intraspecific territoriality among the sylvid warblers of England and Sweden. *Ecological Monographs* 48:351-396.
- Dueser, R. D., and H. H. Shugart, Jr. 1978. Microhabitats in a forest-floor small-mammal fauna. *Ecology* 59:89-98.
- Dueser, R. D., and H. H. Shugart, Jr. 1979. Niche pattern in a forest-floor small-mammal fauna. *Ecology* 60:108-118.
- Green, P. E., and J. D. Carroll. 1976. *Mathematical tools for applied multivariate analysis*. Academic Press, New York, New York, USA.
- Green, R. H. 1971. A multivariate statistical approach to the

- Hutchinsonian niche: bivalve molluscs of central Canada. *Ecology* 52:543-556.
- . 1974. Multivariate niche analysis with temporally varying environmental factors. *Ecology* 55:73-83.
- Harner, E. J., and R. C. Whitmore. 1977. Multivariate measures of niche overlap using discriminant analysis. *Theoretical Population Biology* 12:21-36.
- Hutchinson, G. E. 1957. Concluding remarks. Cold Spring Harbor Symposium in Quantitative Biology 22:415-427.
- . 1968. When are species necessary? Pages 177-186 in R. C. Lewontin, editor. *Population biology and evolution*. Syracuse University Press, Syracuse, New York, USA.
- IMSL. 1979. International mathematical and statistical libraries, Inc. Reference manual. Houston, Texas, USA.
- James, F. C. 1971. Ordinations of habitat relationships among breeding birds. *Wilson Bulletin* 83:215-236.
- Levins, R. 1968. Evolution in changing environments. Princeton University Press, Princeton, New Jersey, USA.
- Maguire, B., Jr. 1973. Niche response structure and the analytical potentials of its relationship to the habitat. *American Naturalist* 107:213-246.
- Maurer, B. A. 1982, in press. Statistical inference for MacArthur-Levins niche overlap. *Ecology* 63.
- M'Closkey, R. T. 1976. Community structure in sympatric rodents. *Ecology* 57:728-739.
- McNaughton, S. J., and L. L. Wolf. 1981. Dominance and niche in ecological systems. *Science* 167:131-139.
- Morrison, D. F. 1967. Multivariate statistical methods. Second edition. McGraw-Hill, New York, New York, USA.
- Rotenberry, J. T., and J. A. Wiens. 1980. Habitat structure, patchiness, and avian communities in North American steppe vegetation: a multivariate analysis. *Ecology* 61:1228-1250.
- Rotenberry, J. T., and J. A. Wiens. 1981. A synthetic approach to principal component analysis of bird/habitat relationships. Pages 197-208 in D. C. Capen, editor. *The use of multivariate statistics in studies of wildlife habitat*. United States Forest Service General Technical Report RM-87, Rocky Mountain Forest and Range Experiment Station, Fort Collins, Colorado, USA.
- Shugart, H. H., Jr., and B. C. Patten. 1972. Niche quantification and the concept of niche pattern. Pages 284-327 in B. C. Patten, editor. *Systems analysis and simulation in ecology*. Volume II. Academic Press, New York, New York, USA.
- Senseth, N. C. 1981. Optimal food selection: some further considerations with special reference to the grazer-hunter distinction. *American Naturalist* 117:457-475.
- Sugihara, G. 1980. Minimum community structure: an explanation of species abundance patterns. *American Naturalist* 116:770-787.
- Tatsuoka, M. M. 1970. Discriminant analysis: the study of group differences. *Selected Topics in Advanced Statistics, An Elementary Approach*. Number 6. Institute for Personality and Ability Testing, Champaign, Illinois, USA.
- Wiens, J. A., and J. T. Rotenberry. 1981. Censusing and the evaluation of avian habitat occupancy. *Studies in Avian Biology* 6:522-532.
- Williams, B. K. 1981. Discriminant analysis in wildlife research: theory and applications. Pages 59-71 in D. C. Capen, editor. *The use of multivariate statistics in studies of wildlife habitat*. United States Forest Service General Technical Report RM-87, Rocky Mountain Forest and Range Experiment Station, Fort Collins, Colorado, USA.

APPENDIX

Proof of equivalence of $\sum d_{ij}^2/n_i$ and $\sum S_{ik}^2$ where n_i = number in i^{th} group, d_{ij} = distance of j^{th} individual from centroid of i^{th} group and S_{ik}^2 is variance of scores of i^{th} group on k^{th} axis.

Let x_{jk} represent the score of the j^{th} individual on the k^{th} axis and \bar{x}_{ik} equal the group mean on the k^{th} axis. Then

$$d_{ij}^2 = \sum_{k=1}^p (x_{jk} - \bar{x}_{ik})^2$$

$$d_{ij}^2/n_i = \sum_{k=1}^p (x_{jk} - \bar{x}_{ik})^2/n_i.$$

Summing over all individuals:

$$\sum_{j=1}^{n_i} d_{ij}^2/n_i = \sum_{j=1}^{n_i} \sum_{k=1}^p (x_{jk} - \bar{x}_{ik})^2/n_i.$$

Reversing the order of summation on the right hand side gives:

$$\sum_{k=1}^p \sum_{j=1}^{n_i} (x_{jk} - \bar{x}_{ik})^2/n_i = \sum_{k=1}^p S_{ik}^2.$$