



Partitioning Mahalanobis D^2 to sharpen GIS classification

J.E. Dunn & L. Duncan

*Department of Mathematical Sciences
University of Arkansas, U.S.A.*

Abstract

Mahalanobis D^2 is in common use to quantify habitat suitability in maps prepared by GIS techniques. This paper demonstrates the utility of partitioning D^2 into a sum of orthogonal components. Geometrically each component is identified as the squared distance, in standard measure, from a plane of closest fit, as originally defined by K. Pearson. Thus, for some small k and any vector measurement, the sum of the k components corresponding to the k smallest, nonzero eigenvalues of the covariance matrix reflects the squared distance of the measurement from the intersection of k hyperplanes in the p -dimensional measurement space. Species requirements, rather than being defined in terms of individual measured variables, are instead defined in terms of combinations of variables which satisfy the equations of these k planes. As a result, species requirements admit to a trade-off among habitat variables so long as overall utility is maintained.

1 Introduction

Increasingly, GIS methodology is used to determine the value of an index for the likelihood of occurrence of a species at each grid point within a study area, based on a multivariate configuration of habitat variables at those points. The resulting maps then depict spatial variation in potential use by that species. References [1] and [2] provide examples, as do several of the citations which they contain. Logistic regression, Bayes discriminant functions, and Mahalanobis D^2 provide



familiar indices of habitat suitability. More recently, Dettmers *et al.* [1] have introduced use of regression trees for classification. However, the first, second and fourth approaches are disadvantaged by the fact that they require two calibration data sets, one being examples of suitable habitat, as determined by the observed presence of the species, and the second representing habitats where the species could not occur. Clearly, if the latter is equated to those habitats where the species has not been observed, as is often the case, then its reliability cannot be guaranteed because of the potential presence of false negatives. By contrast, Clark, Dunn, and Smith [3] pointed out that D^2 utilizes only the first of the two data sets, graduating any potential habitat in terms of its standard measure from the centroid of the target data set of occupied habitats. Based on empirical comparisons of the effectiveness of the four methods, Dettmers, *et al.* [1] remarked "The Mahalanobis distance method ... lacks well-defined procedures for model assessment and determination of significant variables in a given model". This partly motivated the results presented here.

A search for "significant variables" reflects an implicit belief that the species and humans agree on what constitutes "variables". In their expository paper, Rotenberry, Knick and Dunn [4] challenged this, arguing that instead of seeking constancy of variables in species requirements, one should seek constant relationships. Their approach consists of partitioning D^2 into a sum of orthogonal components, identified as squared distances from a collection of hyperplanes, then identifying certain of these components as functional requirements of the species. Section 2 amplifies the mathematical details of this concept. Section 3 explores confirmatory factor analysis as a means of sharpening requirements defined in this way, and the methodology is illustrated with an example drawn from ecology.

In what follows, upper case letters, including Greek, refer to matrices, e.g., S , Σ . Lower case boldface letters are column vectors, e.g., \mathbf{y} , $\boldsymbol{\mu}$. The methodology is described in ecological terms such as species, habitat requirements, etc. because this is our chief area of experience. However, nothing seems to prevent us from identifying observations with customers pushing their shopping carts down the aisles of a typical WalMart Supercenter.

2 Orthogonal decomposition of Mahalanobis D^2

Suppose that p random variables, y_1, \dots, y_p , describe habitat. Define $\mathbf{y} = [y_1, \dots, y_p]'$, and let \mathcal{H} be the set of all habitats which are suitable for the species. Let $\mathcal{E}[\mathbf{y}] = \boldsymbol{\mu}$, $\text{var}[\mathbf{y}] = \Sigma$ for $\mathbf{y} \in \mathcal{H}$. Mahalanobis D^2 , as commonly used in GIS mapping, is defined by

$$D^2 = (\mathbf{y} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu}), \quad (1)$$

interpreted as the squared, standardized distance separating \mathbf{y} and $\boldsymbol{\mu}$. However, if the spectral decomposition (Johnson [5]) of Σ is



$$\Sigma = \sum_{j=1}^p \lambda_j \alpha_j \alpha_j' \quad (2)$$

where $\lambda_1 \geq \dots \geq \lambda_p$ are the eigenvalues of Σ with associated, length one eigenvectors $\alpha_1, \dots, \alpha_p$, then necessarily

$$\Sigma^{-1} = \sum_{j=1}^p \lambda_j^{-1} \alpha_j \alpha_j' \quad (3)$$

This follows since $[\Sigma - \lambda_j I] \alpha_j = \mathbf{0} \Rightarrow [\Sigma^{-1} - 1/\lambda_j] \alpha_j = \mathbf{0}$ for $j = 1, \dots, p$, i.e., eigenvalues of Σ^{-1} are reciprocals of those of Σ and the associated eigenvectors are unchanged. Substitution in (1) yields an orthogonal decomposition of D^2 , namely

$$\begin{aligned} D^2 &= \sum_{j=1}^p (\mathbf{y} - \boldsymbol{\mu})' \alpha_j \alpha_j' (\mathbf{y} - \boldsymbol{\mu}) / \lambda_j \\ &= \sum_{j=1}^p \left(d_j / \sqrt{\lambda_j} \right)^2, \end{aligned} \quad (4)$$

where $d_j = (\mathbf{y} - \boldsymbol{\mu})' \alpha_j$. The task remains to attach meaning to the separate components of D^2 .

From geometry of a linear vector space, d_j represents the length of a projection of "mean-centered" $\mathbf{y} - \boldsymbol{\mu}$ on an axis defined by extensions of α_j (Johnson [5]), where the sign of d_j is determined by the positive orientation of α_j . Since $d_j = 0$ is possible for some \mathbf{y} , implying a projection of length zero on α_j , a crucial point is that α_j is normal to a $p - 1$ dimensional hyperplane defined as all \mathbf{y} satisfying

$$(\mathbf{y} - \boldsymbol{\mu})' \alpha_j = 0. \quad (5)$$

The deviation in the orthogonal sense of $\mathbf{y} - \boldsymbol{\mu}$ from this hyperplane is identical to its projection d_j on the axis defined by α_j , so that the variance of these deviations is given by $\text{var}[d_j] = \alpha_j' \Sigma \alpha_j = \lambda_j \alpha_j' \alpha_j = \lambda_j$ for $\mathbf{y} \in \mathcal{Y}$.

It follows that the "plane of closest fit", in the sense of Pearson [6], corresponds to

$$(\mathbf{y} - \boldsymbol{\mu})' \alpha_1 = 0. \quad (6)$$

since for $\mathbf{y} \in \mathcal{Y}$, deviations $d_1 = (\mathbf{y} - \boldsymbol{\mu})' \alpha_1$ from this plane have the smallest possible variance, namely λ_1 . For any \mathbf{y} , $d_1 / \sqrt{\lambda_1}$ in eqn (4) represents its deviation in standard measure with metric defined by the smallest eigenvalue of Σ . A second-best, $p - 1$ dimensional hyperplane, which satisfies $\text{corr}[d_1, d_2] = 0$, is defined by

$$(\mathbf{y} - \boldsymbol{\mu})' \alpha_2 = 0. \quad (7)$$

with deviations of $\mathbf{y} - \boldsymbol{\mu}$ from this hyperplane reflected by $d_2 = (\mathbf{y} - \boldsymbol{\mu})' \alpha_2$, or $d_2 / \sqrt{\lambda_2}$ in standard measure, and so forth. The net result is that D^2 represents a sum of squares of deviations, in standard measure, of a particular point with



coordinates given by $\mathbf{y} - \boldsymbol{\mu}$ from each of $p, p - 1$ dimensional hyperplanes, all of which pass through the point $\mathbf{y} = \boldsymbol{\mu}$ in the original p -dimensional sample space.

In applications, $\boldsymbol{\mu}$ usually will be replaced by the centroid, $\bar{\mathbf{y}} = n^{-1} \sum_{i=1}^n \mathbf{y}_i$ and Σ by its unbiased estimator, $S = (n - 1)^{-1} \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})'$, where y_1, \dots, y_n are p -dimensional vectors characterizing n habitats known to be occupied by the species. It is informative to explore why, with this substitution, we arrive at the "plane of closest fit" in the orthogonal least squares sense as originally developed by Pearson [6]. If $\boldsymbol{\alpha}$ is any p -dimensional vector normalized to length one, the smallest distance between the point $\mathbf{y}_i - \bar{\mathbf{y}}$ and the hyperplane $\boldsymbol{\alpha}'(\mathbf{y}_i - \bar{\mathbf{y}}) = 0$ is given by the orthogonal projection, $\mathbf{e}_i = \boldsymbol{\alpha}'(\mathbf{y}_i - \bar{\mathbf{y}})$ $i = 1, \dots, n$. The sum of squares of these "errors", computed in a direction normal to the hyperplane, is

$$\begin{aligned} \text{SSE}(\boldsymbol{\alpha}) &= \sum_{i=1}^n \mathbf{e}_i^2 = \sum_{i=1}^n \boldsymbol{\alpha}'(\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})' \boldsymbol{\alpha} \\ &= \boldsymbol{\alpha}' \left\{ \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})' \right\} \boldsymbol{\alpha} \\ &= (n - 1)^{-1} \boldsymbol{\alpha}' S \boldsymbol{\alpha}. \end{aligned} \tag{8}$$

By analogy to a principal components solution, this is minimized by choosing $\boldsymbol{\alpha}$ as the eigenvector of length one corresponding to the minimum eigenvalue of S . But this is what we previously proposed to do, with the understanding that unknown $\boldsymbol{\mu}$ and Σ there would require replacement by their usual estimators. Additional planes of increasingly poorer fit are defined by the eigenvectors associated with an increasing sequence of eigenvalues of S .

2.1 Proposed modification to D^2

Rotenberry, Knick, and Dunn [4] argued the premise that not all p components of D^2 , as partitioned in eqn (4), are likely to define limiting combinations of habitat variables for the species. Some $p - k$ of these are included in D^2 simply because p habitat variables were measured or available in the GIS database. For example, Dettmers, *et al.* [1] initiated their avian habitat analysis with 24 variables. Certainly, the hyperplane $(\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\alpha}_1 = 0$, corresponding to the first principal component, logically cannot be considered a limitation since the variance of deviations from this hyperplane is λ_1 , the maximum possible. Yet, this deviation makes its contribution to D^2 as commonly defined.

As a result, it is proposed that habitat suitability for a p -dimensional \mathbf{y} be measured by

$$D^2(k) = \sum_{j=p-k+1}^p d_j^2 / \lambda_j \tag{9}$$

for some $1 \leq k < p$, where the eigenvalues of Σ (or its sample analog) are ordered $\lambda_1 \geq \dots \geq \lambda_p$. Thus, suitability of a particular habitat location \mathbf{y} for a species would be measured in terms of deviations from k basic requirements for that



species, to the extent that we are able to know k . In that $D^2(k) \sim \chi^2_{(k)}$ under multinormal assumptions, $p = P[\chi^2_{(k)} > D^2(k)]$ is analogous to a posterior probability resulting from use of either a Bayes discriminant function or logistic regression.

To satisfy a criticism by Knick and Rotenberry [2], use of $D^2(k)$ has the potential for coping with a rapidly evolving environment, e.g., rangeland recovering from fire damage, in a way that $D^2 = D^2(p)$ does not. In the latter case, the "ideal" habitat, represented by μ , is a single point at the intersection of p hyperplanes in p -space. Less desirable habitat is inferred by any deviation from μ . For $D^2(k)$, "ideal" habitat corresponds to the locus of the intersection of k hyperplanes in p -space, allowing unlimited species expansion in any of the remaining $p - k$ dimensions without degrading habitat suitability. The only requirement is that habitat variation not proceed in any of the directions parallel to mutually orthogonal axes defined by $\alpha_1, \dots, \alpha_k$.

2.2 Characterization of any plane of closest fit

Insight may be gained by visualizing the habitat constraints implied by each of the components of $D^2(k)$. Collins [7] attempted to identify planes of closest fit by interpreting their correlation structure, mimicking principal components analysis. However, since correlation structure reflects the vagaries of the calibration data whereas we want to infer species requirements given an infinite choice of habitats, it is more informative to relate variation in each of the variables to the rate of departure from the target planes. For the j^{th} component, this information is contained in the gradient vector

$$\frac{\partial d_j}{\partial \mathbf{y}} = \left[\frac{\partial d_j}{\partial y_1}, \dots, \frac{\partial d_j}{\partial y_p} \right] = \alpha_j = \{\alpha_{hj}\}. \quad (10)$$

The rank order of $|\partial d_j / \partial y_h| = |\alpha_{hj}|$ from large to small for $h = 1, \dots, p$ suggests the relative importance of the habitat variables to the species requirement defined by $d_j = 0$. Variables with small $|\alpha_{hj}|$ can vary considerably while allowing the habitat to remain close to the requirement. This description is only complete, however, when one realizes that even if both α_{rj} and α_{sj} are large in absolute value, their effects will tend to cancel if they have opposite signs and the habitat still will satisfy the requirement $d_j \cong 0$. This is a perceived advantage of the model since it allows the possibility of detecting that a species can make a trade-off, balancing different habitat variables, while still maintaining habitat utility.

For the above interpretation to be effective, y_1, \dots, y_p should be in identical units. An obvious approach is to standardize all variables preliminary to the habitat analysis, i.e., replacing \mathbf{y} with $\mathbf{y}_s = D_\sigma^{-1}(\mathbf{y} - \mu)$, where diagonal D_σ



displays the standard deviations of the elements of \mathbf{y} . The net effect is to replace Σ by \mathbf{R} , the correlation matrix. While this has no effect on $D^2(\mathbf{p})$ since

$$D^2(\mathbf{p}) = \mathbf{y}'_s \mathbf{R}^{-1} \mathbf{y}_s = (\mathbf{y} - \boldsymbol{\mu})' \mathbf{D}_\sigma^{-1} (\mathbf{D}_\sigma^{-1} \Sigma \mathbf{D}_\sigma^{-1}) \mathbf{D}_\sigma^{-1} (\mathbf{y} - \boldsymbol{\mu}) \\ = (\mathbf{y} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu}),$$

individual components will differ. This is traceable to the nonlinear relationships between polynomial roots of $|\Sigma - \lambda \mathbf{I}| = 0$ and those of the generalized eigenvalue problem $0 = |\mathbf{R} - \theta \mathbf{I}| = |\mathbf{D}_\sigma^{-1} \Sigma \mathbf{D}_\sigma^{-1} - \theta \mathbf{I}| = |\Sigma - \theta \mathbf{D}_{\sigma^2}|$.

3 Variable selection

Choice of k often is compounded by a perceived need to reduce the number of habitat variables involved in the defining relationships. Suppose a tentative k is chosen on the basis of an initial principal components solution, $\Sigma = \Lambda \Lambda'$, and we wish to test (after some rearrangement of variables)

$$H_0: \Lambda = \begin{bmatrix} \overset{(p-k)}{A_{11}} \mathbf{D}_{\sqrt{\{s1\}}} & \overset{(k)}{A_{12}} \mathbf{D}_{\sqrt{\{s2\}}} \\ \overset{(p-k)}{A_{21}} \mathbf{D}_{\sqrt{\{s1\}}} & \text{(all zero)} \end{bmatrix} \begin{matrix} (r) \\ (p-r) \end{matrix} \quad (11)$$

i.e., that k species requirements depend only on r habitat variables, where $\mathbf{D}_{\sqrt{\{s1\}}} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_{p-k}})$, $\mathbf{D}_{\sqrt{\{s2\}}} = \text{diag}(\sqrt{\lambda_{p-k+1}}, \dots, \sqrt{\lambda_p})$, while

$[A'_{11}, A'_{21}]' = [\alpha_1, \dots, \alpha_{p-k}]$ and $[A'_{12}, \text{(all zero)}]' = [\alpha_{p-k+1}, \dots, \alpha_p]$ contain the eigenvectors of Σ as columns. The solution $\Sigma = \Lambda \Lambda'$ corresponds to

$$\Sigma = \begin{bmatrix} \overset{(r)}{\Sigma_{11}} & \bullet \\ \dots & \dots \\ \bullet & \bullet \end{bmatrix} \begin{matrix} (r) \\ \dots \\ (r) \end{matrix} = \begin{bmatrix} \overset{(r)}{A_{11} \mathbf{D}_{\{s1\}} A'_{11} + A_{12} \mathbf{D}_{\{s2\}} A'_{12}} & \bullet \\ \dots & \dots \\ \bullet & \bullet \end{bmatrix} \begin{matrix} (r) \\ \dots \\ (r) \end{matrix}.$$

But when H_0 is true, $(A_{11} \mathbf{D}_{\{s1\}} A'_{11} + A_{12} \mathbf{D}_{\{s2\}} A'_{12}) A_{12} = A_{12} \mathbf{D}_{\{s2\}}$ since

$A'_{11} A_{12} = \mathbf{0}$ and $A'_{12} A_{12} = \mathbf{I}_{(k)}$ from the orthogonality requirements placed on the original solution. Thus, under H_0 , the original eigenvectors defining k habitat requirements in eqn (11) are contained in the set of eigenvectors of Σ_{11} involving only r habitat variables. Since $k \leq r$, the required eigenvectors correspond to the k smallest eigenvalues of Σ_{11} (and Σ).

In our consulting experience, the proposed sequence of tests has produced a sharp demarcation between zero and non-zero component loadings. The



following is an illustration taken from Gullett [8] in which the calibration data set consisted of 97 pixels in the Ouachita mountains of Arkansas where the Acadian flycatcher had been observed. Each pixel was characterized by 11 logratios describing forest species composition at the site. An initial principal components solution produced eigenvalues $\lambda_1 = 4.273, \dots, \lambda_8 = 0.127, \lambda_9 = 0.097, \lambda_{10} = 0.069, \lambda_{11} = 0.024$, so that $k = 3$ was chosen on the basis of subjective judgment. The last three eigenvectors are reproduced below:

	(α_9)	(α_{10})	(α_{11})
LR1	-0.54488 (3)	0.00640 (2)	0.00364 (2)
LR2	0.10296 (2)	0.00656 (2)	0.78384 (8)
LR4	0.75496 (7)	0.02664 (2)	0.07768 (2)
LR5	0.15504 (2)	-0.18829 (2)	-0.56979 (4)
LR6	-0.01580 (1)	-0.06135 (1)	0.03411 (1)
LR10	0.18769 (1)	0.19772 (1)	-0.20949 (1)
LR11	-0.20557 (1)	0.09496 (1)	0.05170 (1)
LR12	0.03453 (2)	0.68456 (5)	-0.06051 (2)
LR17	-0.04113 (1)	-0.03171 (1)	0.02506 (1)
LR18	0.05763 (2)	-0.66277 (6)	0.05194 (2)
LR20	-0.12066 (1)	-0.05464 (1)	-0.01321 (1)

Numbers in parenthesis indicate the order in which confirmatory tests were performed using SAS[®] procedure CALIS [9], with all tests based on 55 degrees of freedom. Step (1) attempted to zero entire rows LR6, LR10, LR11, LR17, and LR20, yielding $X^2 = 12.44$ with $p = 1.0$. This suggests that these 5 logratios do not contribute to the 3 habitat requirements of the species. In addition to the zero elements specified in step (1), the hypothesis of step (2) also specified certain zero elements in those rows containing at least one sizable element, e.g., LR1. Again, this hypothesis was accepted since $X^2 = 15.70, p = 1.0$. Steps (3) - (8) included the hypotheses of steps (1) and (2) as well as zeroing one additional element. $X^2 = 114.8, 96.3, 260.9, 213.9, 821.9, \text{ and } 588.8$, respectively, in steps (3) - (8) with $p \leq 0.0005$ in every case, thus leading to retention of the 6 boldface entries shown above. A principal components analysis of the retained variables, LR1, LR2, LR4, LR5, LR12, and LR18, yielded eigenvalues $\lambda_4 = 0.137, \lambda_5 = 0.080$, and $\lambda_6 = 0.031$ with associated eigenvectors

	(α_4)	(α_5)	(α_6)
LR1	-0.57824	0.08683	-0.04937
LR2	0.01196	0.00277	0.73950
LR4	0.77417	0.00837	0.15787
LR5	0.24422	0.03844	-0.65173
LR12	-0.07531	-0.72059	-0.00618
LR18	-0.02907	0.68677	0.03133



202 *Management Information Systems*

Aside from an immaterial reflection of signs in one case, the eigenvalues and eigenvectors for the reduced set of 6 variables closely mimic those of the original set of 11 variables. These three eigenvectors ultimately were the basis for $D^2(3)$ used to map potential habitat of the Acadian flycatcher. The boldface elements suggest that Acadian flycatcher habitat requirements will be met if (a) all 6 logratios are close to their respective means, or if (b) LR1 and LR4 simultaneously vary in the same direction from their means, or if (c) LR12 and LR18 simultaneously vary in the same direction from their means, or if (d) LR2 and LR5 simultaneously vary in the same direction from their means, or if (e) any combination of (b), (c), or (d) holds.

Figure 1 shows an overlay of the empirical c.d.f. of $D^2(3)$ for the 97 known Acadian flycatcher sites, its kernel density smooth by means of SAS[®] procedure KDE [9], and the c.d.f. of the approximating $\chi^2_{(3)}$. We used 51 bins and a bandwidth multiplier of 1.3 for the kernel density estimation shown here. It is typical that the empirical c.d.f. tends to accumulate faster than the normal-based chisquared distribution, suggesting that one or both tails of the variables defining $D^2(3)$ have been trimmed in the calibration data set. This corresponds to preferred habitat being in the middle of the ranges of the habitat variables, a situation where both Bayes linear discriminant functions and logistic regression tend to perform rather poorly.

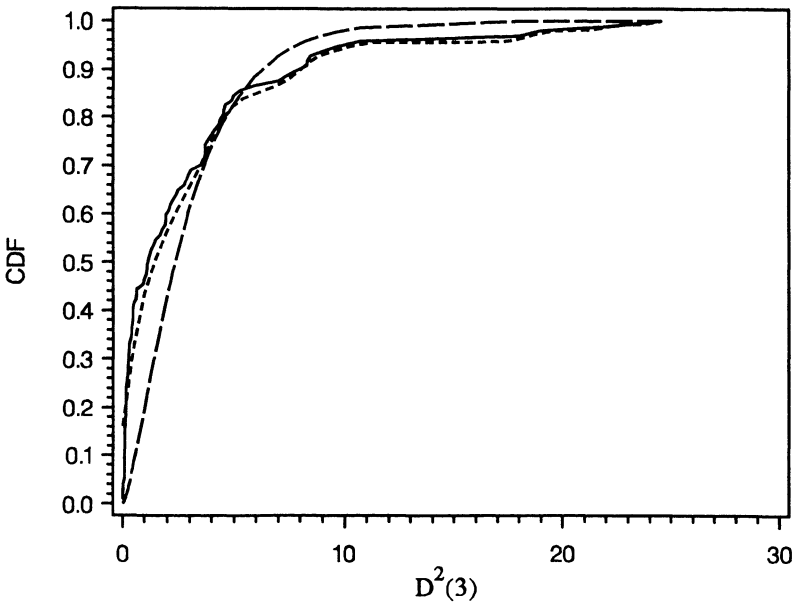


Figure 1: Empirical (—) and smoothed (-----) c.d.f.'s for $D^2(3)$ at 97 flycatcher sites, and cumulative $\chi^2_{(3)}$ (-.-).



4 Discussion

Objections may be leveled at the subjective nature of the hypotheses being tested above in arriving at a relevant subset of habitat variables. This is a flaw inherent in all stepwise model building algorithms. The practicality of the results are our only defense. Choice of k is likely to be somewhat subjective, depending on the magnitudes and relative spacings among the eigenvalues, the interpretability of $d_{p-k+1} = 0, \dots, d_p = 0$, and the credibility of predicted use areas which result from particular choices of k . In this respect, it does not differ from any other principal components/factor analysis application where interpretability usually dictates choice of the number of dimensions. We find it surprising that the empirical c.d.f., or its smoothed version, has not entered into general use for assignment of p -values to D^2 in the context of GIS habitat mapping. It is an obvious answer to concerns about non-normality often encountered. Perhaps by tradition D^2 and its asymptotic chisquared distribution are irrevocably linked, but this is a misconception. D^2 and $D^2(k)$ are valid metrics, requiring only second moment properties, regardless of what distributions are associated with them.

Of the variety of problems which we have encountered in using $D^2(k)$, the most pervasive has been the presence of a zero eigenvalue in the initial principal components solution. Interpreted literally, this implies a mechanistic “all-or-nothing” species requirement corresponding to the hyperplane defined by the associated eigenvector. Such might be the case for habitats supporting an avian species, say, if such universal requirements as the presence of water, food supply, nesting sites, etc. were included as indicator variables in the data set. However, the situation corresponding to mutual absence of all measurable properties is not so easily understood. Our problems always have arisen in situations where the data takes the form

(y_r)	(y_s)	(y_t)
.	.	.
0	0	0
a	0	c
0	b	d
0	0	0
.	.	.

This situation typically arises when focusing on a subregion within a vegetatively heterogeneous data base. Even though y_r, y_s, y_t , are quantitative variables, the rarity of their occurrence causes them to behave as indicator variables and linear column dependencies readily appear. To define a species requirement as the mutual absence of measurable properties seems presumptuous. But still, as humans, our own preferred habitat is determined by the (near) absence of UV radiation, absence of pollution, etc. The problem ultimately comes down to asking what level of subdivision of the microhabitat is biologically meaningful, so that further refinements are merely bothersome.



References

- [1] Dettmers, R., Buehler, D.A. & Bartlett, J.G. A test and comparison of wildlife-habitat modeling techniques for predicting occurrence on a regional scale. *Predicting Species Occurrence: Issues of Scale and Accuracy*, conference October 18-22, 1999, Snowbird, Utah.
- [2] Knick, S.T. & Rotenberry, J.T. Limitations to mapping habitat use areas in changing landscapes using Mahalanobis distance statistic. *Journal of Agricultural, Biological and Environmental Statistics*, **3**, pp. 311-322, 1998.
- [3] Clark, J.D., Dunn, J.E. & Smith, K.G. A multivariate model of female black bear habitat use for a geographic information system. *Journal of Wildlife Management*, **57(3)**, pp. 519-526, 1993.
- [4] Rotenberry, J.T., Knick, S.T. & Dunn, J.E. A minimalist approach to mapping species habitat: Pearson's planes of closest fit. *Predicting Species Occurrence: Issues of Scale and Accuracy*, conference October 18-22, 1999, Snowbird, Utah.
- [5] Johnson, D.E. Distances and Angles (Appendices A.3 & A.4). *Applied Multivariate Methods for Data Analysis*, Duxbury Press: London and New York, pp. 528-529, 1998.
- [6] Pearson, K. On lines and planes of closest fit in systems of points in space. *Philosophical Magazine*, **2**, pp. 559-572, 1901.
- [7] Collins, S.L. Geographic variation in habitat structure of the black-throated green warbler (*Dendroica virens*). *The Auk*, **100**, pp. 382-389, 1983.
- [8] Gullett, B.W. Combining Arkansas breeding bird atlas and GAP vegetation data to predict avian distributions within the Ouachita Mt. Physiographic Region. Thesis submitted to the University of Arkansas, Dept. of Biological Sciences, 2000.
- [9] *SAS OnlineDoc Version 8*. SAS Institute, Inc., Cary, NC, 2000.