

## Classifying herbivore diets using hierarchical cluster analysis

MICHAEL L. MCINNIS, LARRY L. LARSON, AND MARTIN VAVRA

### Abstract

Hypotheses are frequently posed that require comparisons among diet samples. Hierarchical cluster analysis is suited to this task, but has received little attention in food habits research. By grouping samples so that similar diets are close together and dissimilar diets are spaced farther apart, hierarchical cluster analysis reveals patterns in data difficult to recognize in the matrix of similarity coefficients typical of most food habits studies. We provide an example of this type of analysis, and indicate its application to management of large herbivores.

**Key Words:** classification, diet determination techniques, cluster analysis, hierarchical cluster analysis, cattle diets, horse diets, pronghorn diets

Frequently, hypotheses are posed that require comparison of diet samples to one another. This usually involves calculating resemblance of diets among animal species within specified time intervals, or within species among time intervals. Typically, resemblance coefficients (similarity or dissimilarity indices) are calculated between pairs of samples, and results are tabulated in a two-way matrix with 1 coefficient for each pair of samples (Holechek et al. 1984). By itself, this "resemblance matrix" may not clearly reveal the structure of large data sets because groups of similar samples are not identified (Romesburg 1984). Grouping samples is advantageous because it refines large data sets by identifying commonality among a number of samples, and uniqueness in samples that do not fit a particular group (Gauch 1982). This is desirable in food habits research because, when integrated with such information as availability or nutritional quality of food items, grouping diet samples can provide management insights into feeding strategies or exploitative competition (Hanley 1982).

Samples can be grouped by nonhierarchical or hierarchical classification (Williams 1971). Nonhierarchical methods identify groups (clusters) of similar samples but do not characterize relationships among clusters (Gauch 1982). Hierarchical classification identifies clusters of similar samples and, additionally, "arranges the groups into a hierarchical, treelike structure called a dendrogram, which indicates relationships among the groups" (Gauch 1982).

Hierarchical cluster analysis has received little attention in food habits research. This paper describes the use of hierarchical cluster analysis as a method for classifying and interpreting herbivore diets. As an example, we classified the diets of sympatric feral horses (*Equus caballus*), cattle (*Bos taurus*), and pronghorn (*Antilocapra americana*), using the data of McInnis and Vavra (1987).

### Methods

McInnis and Vavra (1987), in a study of forage partitioning among sympatric feral horses, cattle, and pronghorn, used fecal analysis to determine seasonal diets of these herbivores from May

1979 through March 1981. Five fecal samples per month were prepared for each species of herbivore. Each sample was a composite of feces collected from 6–12 observed individuals. Monthly means were expressed as percent relative density and pooled to yield 3 replications per season (spring = March–May; summer = June–August; fall = September–November; winter = December–February).

In the present study, we used hierarchical cluster analysis to classify mean seasonal diets of these herbivores. Romesburg (1984) outlined the steps involved in cluster analysis: (1) obtain the data matrix; (2) compute the resemblance matrix; (3) execute the clustering method; and (4) compute the cophenetic correlation coefficient. The data matrix used in our sample consisted of the mean percent composition of forage species in the seasonal diets of each herbivore species (McInnis and Vavra 1987). Similarity of botanical composition between each pair of diets was calculated using Romesburg's (1984) modification of the Bray-Curtis resemblance coefficient:

$$b_{jk} = 100 \frac{\sum_{i=1}^n |x_{ij} - x_{ik}|}{\sum_{i=1}^n (x_{ij} + x_{ik})}$$

where  $b_{jk}$  was the percent similarity between the diets of herbivores  $j$  and  $k$ , and may range from 0–100%;  $n$  was the number of forage species in 2 diets; and  $x_{ij}$  and  $x_{ik}$  were proportions of forage species  $x_i$  in the diets of herbivores  $j$  and  $k$ , respectively. Numerous clustering methods are available, and have been reviewed by Sneath and Sokal (1973), Clifford and Stephenson (1975), and Romesburg (1984). Williams (1971) classified hierarchical clustering methods as divisive or agglomerative. Divisive clustering results in the analysis of an entire population of samples into increasingly finer partitions, whereas agglomerative methods group individual samples into increasingly coarser partitions, arriving eventually at a single set containing all samples (Sneath and Sokal 1973). Agglomerative methods dominate published uses because they were made popular by early literature and have been the object of empirical analysis (Blashfield and Aldenderfer 1978). In the present study, complete linkage (also called "furthest neighbor") clustering (Sneath and Sokal 1973, Romesburg 1984) of mean seasonal diets was accomplished using the computer program CLUSTER (Keniston 1978). Complete linkage is an agglomerative clustering strategy that computes similarity of 2 clusters as the most dissimilar pair of objects, 1 in each cluster (Romesburg 1984). Results of the cluster analysis were summarized as a dendrogram ("tree"). Statistical differences between diets were determined using the  $t$ -test for unpaired observations and unequal variances (Steel and Torrie 1960) to examine the accuracy of the classification. This was accomplished using mean monthly composition of individual forage species as observations. The cophenetic correlation coefficient ( $r_{xy}$ ) was calculated to evaluate distortion between the dendrogram and the resemblance matrix (Romesburg 1984).

### Results and Discussion

Typical food habits studies include a primary data matrix show-

McInnis and Larson are assistant professors, Department of Rangeland Resources, Oregon State University, Corvallis, and are stationed at OSU-EOSC Agriculture Program, Eastern Oregon State College La Grande, 97850; Vavra is professor, Department of Rangeland Resources, and superintendent, Eastern Oregon Agricultural Research Center, Rt. 1, Box 4.51 Hwy. 205, Burns, 97220.

Published with approval of the Director, Oregon Agricultural Experiment Station, as Technical Paper 8473.

Manuscript accepted 20 September 1989.

Table 1. Spring-summer vs. fall-winter diets of feral horses and cattle.

Forage class Forage species	Spring-Summer			Fall-Winter		
	% Dietary Composition (±SE)		Significance <sup>1</sup>	% Dietary Composition (±SE)		Significance <sup>1</sup>
	spring-summer	fall-winter		spring-summer	fall-winter	
	n = 9	n = 12		n = 9	n = 12	
Total grasses	92.0(1.9)	86.1(1.6)	+	91.7(0.9)	86.6(1.4)	+
<i>Agropyron spicatum</i>	26.6(3.1)	9.6(1.1)	**	31.5(4.3)	11.6(1.4)	**
<i>Sitanion hystrix</i>	20.4(3.9)	43.1(2.7)	**	18.6(3.9)	56.6(1.9)	**
<i>Bromus tectorum</i>	13.3(2.9)	3.1(0.6)	**	13.9(2.2)	5.6(1.0)	**
<i>Poa sandbergii</i>	5.0(1.9)	11.9(2.4)	+	8.5(1.4)	3.1(0.6)	**
<i>Stipa thurberiana</i>	20.5(2.4)	12.9(1.8)	+	4.9(0.8)	4.0(1.4)	ns
Total forbs	4.8(0.9)	9.9(1.4)	+	7.3(0.9)	1.9(0.4)	**
Total shrubs	3.2(1.2)	4.0(0.8)	ns	1.0(0.4)	11.4(1.5)	**
<i>Atriplex confertifolia</i>				0.5(0.3)	10.0(1.6)	**

<sup>1</sup>+, \*\* = significant at 0.10 and 0.01 levels, respectively; ns = nonsignificant ( $p > 0.10$ ).

ing botanical composition of each diet, a resemblance matrix indicating similarity among diets, and results of statistical analyses comparing differences of these diets (Holechek et al. 1982, Krysl et al. 1984). When the number of samples is small, this approach may satisfactorily reveal relationships among diets. Holechek et al. (1984) described the use of multivariate analysis of variance and a multiple-range test to compare each diet with all other diets when the F-test is significant. The authors recommended a resemblance matrix be constructed and "an asterisk used to denote diet pairs differing significantly". Such an approach may be suitable for testing hypotheses when dealing with 5 or fewer variables, but may fail to reveal adequately the structure of larger (two-way matrices of at least  $10 \times 10$ ) data sets (Gauch 1982). In the present study, 21 diet samples were considered (3 herbivore species, 7 seasons) yielding 210 resemblance coefficients. Here, the resemblance matrix alone cannot adequately indicate similarities and differences among the studied diets. When dealing with large data sets, multivariate methods of classification such as hierarchical cluster analysis are advantageous over pair-wise statistical comparison of diets because groups of similar diets are formed. Such groups are especially useful in eliciting the internal structure of large data sets.

The dendrogram resulting from cluster analysis orders diets in a hierarchical manner, with similar diets close together and dissimilar diets spaced farther apart (Fig. 1). Thus, the dendrogram clearly shows patterns in the data difficult to recognize in the resemblance matrix. In the dendrogram, diets may be compared at any level of similarity. The level at which the tree is cut determines the number of clusters formed (Romesburg 1984). For example, when the tree is cut in the range  $b_{jk} < 4$ , only 1 cluster contains all 21 diet samples; when the cut is made so that  $b_{jk} > 89$ , each diet sample forms its own cluster. Clearly, both choices are unsatisfactory in defining relationships among diets of these herbivores, and some number of groups between 1 and 21 must be identified.

The decision of where to cut the tree to form groups is subjective in that no criteria are used to determine when significant changes occur in percent similarity. A practical rule is to cut the tree at a level of  $b_{jk}$  to produce clusters maximally related to variables of interest, and in a manner such that samples are relatively constant within clusters (Romesburg 1984). In our example, variables of interest are herbivore species and season, and our classification of diet samples should relate to each.

The literature suggests ways in which seasonal diets of horses, cattle and pronghorn should be classified. Schoener (1974) found dietary specialization to be a major method of resource partitioning within animal communities. Reasons for dietary partitioning among ungulates have been explained on the basis of body size, type of digestive system (cecal or ruminant), rumino-reticular volume, and mouth size (Hanley 1982). This model predicts large

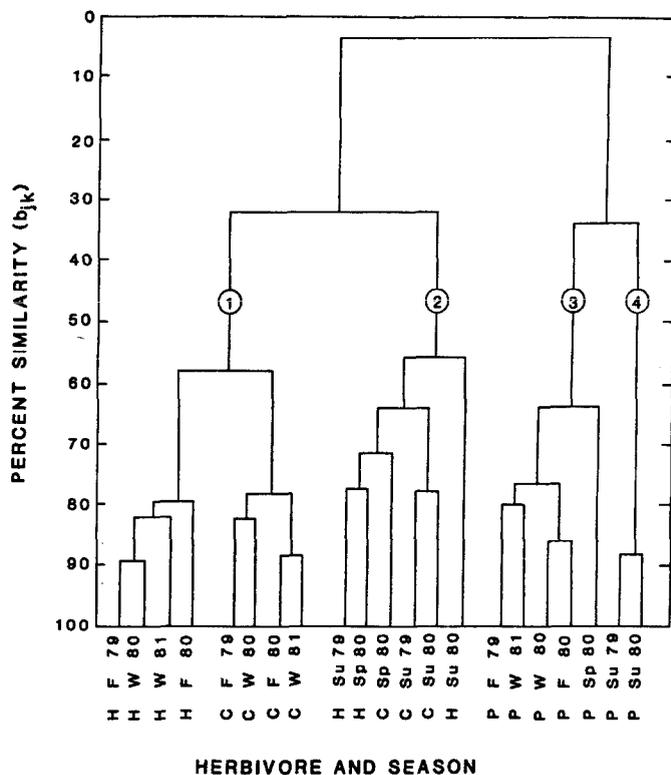


Fig. 1. Dendrogram resulting from cluster analysis of seasonal diets of feral horses (H), cattle (C), and pronghorn (P). F = fall; W = winter; Su = summer; Sp = spring. Circled numbers indicate final diet clusters.

cecal digestors (horses) and small ruminants (pronghorn) be positioned at opposite ends of a less selective graminoid diet to a highly selective forb and browse diet gradient, respectively, with cattle positioned nearer to the horse than to the pronghorn. Further, ecologically similar species may coexist by partitioning resources among themselves (Schoener 1974), particularly during the lean season (Smith et al. 1978). Thus, we expected horses and cattle to be ecologically similar and consume more similar graminoid diets than pronghorn; and diets of horses and cattle to be less similar in fall and winter than other seasons of the year, assuming these are the "lean seasons" (periods of low forage availability and quality). We used the dendrogram resulting from cluster analysis to examine these hypotheses by classifying our diet samples.

Four clusters were formed by cutting the tree in the range  $34 < b_{jk} < 56$  (Fig. 1). Cutting the tree in this range provides a fairly great width in  $b_{jk}$  (22 units) so that botanical composition of diets

**Table 2. Forage species in the diets of feral horses vs. cattle during spring-summer and fall-winter periods.**

Forage class Forage species	Feral Horses		Significance <sup>1</sup>	Cattle		Significance <sup>1</sup>
	% Dietary Composition (±SE)			% Dietary Composition (±SE)		
	Feral Horses	Cattle		Feral Horses	Cattle	
	n = 9	n = 9		n = 12	n = 12	
Total grasses	92.0(1.9)	91.7(0.9)	ns	86.1(1.6)	86.6(1.4)	ns
<i>Agropyron spicatum</i>	26.6(3.1)	31.5(4.3)	ns	9.6(1.1)	11.6(1.4)	ns
<i>Sitanion hystrix</i>	20.4(3.9)	18.6(3.9)	ns	43.1(2.7)	56.6(1.9)	**
<i>Bromus tectorum</i>	13.3(2.9)	13.9(2.2)	ns	3.1(0.6)	5.6(0.1)	+
<i>Poa sandbergii</i>	5.0(1.9)	8.5(1.4)	ns	11.9(2.4)	3.1(0.6)	**
<i>Stipa thurberiana</i>	20.5(2.4)	4.9(0.8)	**	12.9(1.8)	4.0(1.5)	**
Total forbs	4.8(0.9)	7.3(0.9)	+	9.9(1.4)	2.0(0.4)	**
Total shrubs	3.2(1.3)	1.0(0.4)	ns	4.0(0.8)	11.4(1.5)	**

<sup>1</sup>ns = nonsignificant ( $p > 0.10$ ); \*\*, + = significant at 0.01 and 0.10 levels, respectively.

within clusters is relatively uniform. Clusters 1 and 2 contain all horse and cattle diets and are combined at  $b_{jk} = 32$ . Clusters 3 and 4 contain all pronghorn diets and are joined at  $b_{jk} = 34$ . These 2 larger clusters are combined into a single group at  $b_{jk} = 4$ . The reason for the distinction between these larger groups is clear; horses and cattle consumed diets composed principally of grasses, while pronghorn diets were dominated by forbs and shrubs (McInnis and Vavra 1987). This result supports our initial hypothesis.

Clusters 1 and 2 are separated on the basis of season. Cluster 1 contains 2 distinct subclusters: 1 combines the fall/winter diets of horses, and a second joins the fall/winter diets of cattle. Cluster 2 contains the collective spring/summer diets of horses and cattle without clear subclusters. The distinction between clusters 1 and 2 resulted from significant shifts in diets of horses and cattle from spring/summer to fall/winter (Table 1). For horses, use of bottlebrush squirreltail (*Sitanion hystrix* (Nutt.) J.G. Smith) and Sandberg bluegrass (*Poa sandbergii* Vasey) increased during fall/winter compared to spring/summer, while bearded bluebunch wheatgrass (*Agropyron spicatum* (Pursh) Scribn. & Smith) and cheatgrass (*Bromus tectorum* L.) decreased. For cattle, consumption of bottlebrush squirreltail and shadscale saltbrush (*Atriplex confertifolia* (Torr. and Frem.) Wats.) was greater in the fall/winter period, but use of total grasses, bearded bluebunch wheatgrass, Sandberg bluegrass, and total forbs was less.

Spring and summer are generally periods of adequate forage availability on western rangelands compared to fall and winter. Assuming the hypothesis of Smith et al. (1978) is correct, we expected greater dietary divergence between horses and cattle during fall/winter months than during spring/summer. Cluster 1 shows distinct separation between horse and cattle diets in fall/winter, while cluster 2 indicates these 2 herbivores consumed diets not easily distinguishable during spring/summer (Fig. 1). Results of *t*-tests comparing diets of these ungulates during the 2 periods demonstrate few differences in the botanical composition of horse and cattle diets during spring/summer compared to fall/winter (Table 2), thus supporting our hypothesis of greater dietary partitioning among ecologically similar species during lean seasons.

Pronghorn diets are separated by season in clusters 3 and 4. The collective fall/winter diets of pronghorn are joined in cluster 3 at  $b_{jk} = 76$ . This resulted from the preponderance of sagebrush (*Artemisia* spp.) in these diets compared to spring/summer (McInnis and Vavra 1987). The spring diet is intermediate in composition of forbs and browse between fall/winter and summer. Therefore, the spring diet is intermediate to fall/winter and summer in the dendrogram (Fig. 1), and is joined to the fall/winter subcluster at  $b_{jk} = 63$ . Pronghorn switched to diets dominated by forbs and grasses during summer (McInnis and Vavra 1987). The dendrogram reflects this change by grouping summer diets in cluster 4.

Because the clustering procedure may distort some values of the

resemblance coefficients in the dendrogram compared to the resemblance matrix, a method of assessing congruence between the 2 is necessary. The cophenetic correlation coefficient is the measure used most often (Romesburg 1984). It is simply the Pearson product-moment correlation coefficient ( $r_{x,y}$ ) between corresponding values of  $b_{jk}$  taken from the resemblance matrix and the dendrogram. For example, the value of  $b_{jk}$  between cattle diets and horse diets during summer 1979 is 72 in the resemblance matrix. The value of these 2 diets as determined from the dendrogram as  $b_{jk} = 63$  (Fig. 1).

Corresponding values of  $b_{jk}$ , thus, are determined for all other pairs of diets in the resemblance matrix and dendrogram, and  $r_{x,y}$  is calculated. In our example  $r_{x,y} = 0.94$ . This is less than perfect correlation ( $r_{x,y} = 1.0$ ), but well above the point of no correlation ( $r_{x,y} = 0.0$ ). While exact guidelines cannot be defined for the amount of distortion tolerable, values of  $r_{x,y} \geq 0.80$  indicate distortion is not great (Romesburg 1984).

Numerous computer programs are available for cluster analysis. Some are included in such popular statistical analysis packages as SAS, BMDP, SPSS, and CLUSTAN. Others, such as NTSYS, CLUSTAR, and CLUSTID, are specialized programs designed for research in numerical taxonomy. Romesburg (1984) presented a description and critique of these software packages.

### Conclusions

Hierarchical cluster analysis is a general term for mathematical methods that show relationships among objects at increasing levels of similarity. It is a useful tool for classifying food habits data because it produces a dendrogram that clearly emphasizes similarities and differences in kinds and amounts of forages consumed by herbivores within and among seasons. Separation of clusters in the dendrogram is subjective but should produce groups maximally related to variables of interest and maintain uniformity of samples within clusters. Distortion of resemblance coefficients in the dendrogram compared to the resemblance matrix can be assessed using the cophenetic correlation coefficient. As our example demonstrates, classifying herbivore diets using hierarchical cluster analysis can be used to generate or examine hypotheses useful to management.

### Literature Cited

- Blashfield, R.K., and M.S. Aldenderfer. 1978. The literature on cluster analysis. *Multivar. Behav. Res.* 13:271-295.
- Clifford, H.T., and W. Stephenson. 1975. An introduction to numerical classification. Academic Press, New York.
- Gauch, H.G. 1982. *Multivariate analysis in community ecology*. Cambridge Univ. Press.
- Hanley, T.A. 1982. The nutritional basis for food selection by ungulates. *J. Range Manage.* 35:146-151.

**Holechek, J.L., M. Vavra, J. Skovlin, and W.C. Krueger. 1982.** Cattle diets in the Blue Mountains of Oregon, I. Grasslands. *J. Range Manage.* 35:109-112.

**Holechek, J.L., M. Vavra, and R.D. Pieper. 1984.** Methods for determining the botanical composition, similarity, and overlap of range herbivore diets, p. 425-471 *In: National Research Council/National Academy of Science, Developing strategies for rangeland management.* Westview Press, Boulder, Colo.

**Keniston, J.A. 1978.** Program CLUSTER: an aid to numerical classification using CDC Cyber computer. Pleuronectid Proj. Tech. Rep. 2, Oregon State Univ., Sea Grant, Corvallis.

**Krysl, L.J., M.E. Hubbert, B.F. Sowell, G.D. Plumb, T.K. Jewett, M.A. Smith, and J.W. Waggoner. 1984.** Horses and cattle grazing in the Wyoming Red Desert, I. Food habits and dietary overlap. *J. Range Manage.* 37:72-76.

**McInnis, M.L., and M. Vavra. 1987.** Dietary relationships among feral horses, cattle, and pronghorn in southeastern Oregon. *J. Range Manage.* 40:60-66.

**Romesburg, H.C. 1984.** Cluster analysis for researchers. Lifetime Learning Publ., Belmont, Calif.

**Schoener, T.W. 1974.** Resource partitioning in ecological communities. *Science* 185:27-39.

**Smith, J.N.M., P.R. Grant, I.J. Abbott, and L.K. Abbott. 1978.** Seasonal variation in feeding habits of Darwin's ground finches. *Ecology* 59:1137-1150.

**Sneath, P.H.A., and R.R. Sokal. 1973.** Numerical taxonomy. W.H. Freeman and Co., San Francisco.

**Steel, R.G.D., and J.H. Torrie. 1960.** Principles and procedures of statistics. McGraw-Hill Book Co., New York.

**Williams, W.T. 1971.** Principles of clustering. *Annu. Rev. Ecol. Syst.* 2:303-326.