COENOSES 7(3):161-167, 1992

TESTING FOR ELLIPTICAL CLUSTERS IN ECOLOGICAL MULTIDIMENSIONAL SPACES

E. Feoli and M. Lagonegro

Dipartimento di Biologia, Università degli Studi di Trieste, 34100 Trieste, Italia

Keywords: Clusters, Ellipsoids, Multidimensional space, Niche

Abstract. The use of ellipsoids of equal concentration to allocate clusters and to calculate hypervolumes and overlap of niches in ecological multidimensional spaces is suggested.The utility and limitations of the technique are stressed. Program TANTIT is offered for computations.

Introduction

There are no theoretical reasons to assign a priori shape to clusters. If the relationship between the variables within a cluster is linear then the the cluster assumes an ellipsoidal shape (Orlóci 1978). Clusters in ecological spaces may be detected by many methods of clustering (Anderberg 1973, Orlóci 1978, Legendre and Legendre 1983), however statistical testing in general is confined to a handful of methods such as those in the group of discriminant analysis for quantitative data (Mardia, Kent and Bibby 1979), or information analysis for qualitative data (Orlóci 1978, Feoli, Lagonegro and Orlóci 1984), or the probabilistic methods of Goodall (1964, 1966, 1968) and Goodall and Feoli (1988) for mixed data. In traditional statistics, testing amounts to finding a probability of accepting some value being more extreme than some other value, or a mathematical function being suitable for fitting a relationship between two or more sets. In this paper testing does not mean to verify a hypothesis but to measure how much the ellipsoids constructed around the centroids of the clusters are efficient for allocating the regions of the clusters into multidimensional spaces.This exercise is useful to test if the classification chosen is efficient in defining clusters in which the relationship between the variables is linear. It is also useful to measure hypervolumes of ecological niches of species or communities in a more precise way than previuosly (Feoli, Ganis and Zerihun 1988) if the niche has an ellipsoidal shape. The use of ellipses of equal concentration for interpreting ordination patterns in community ecology was already suggested by Lagonegro and Feoli(1985). However, notwithstanding its considerable potentials, the technique is still neglected and remains confined to statistical texts that are not falmiliar to ecologists

In the present paper the technique is formulated assuming spaces of more than two dimensions. A code for computation, TANTIT, has been written by M. Lagonegro, the program has been inserted in the package FIVEPA (Feoli, Lagonegro and Orlóci 1990).

Method

We are considering n clusters and we use k variables (out of a set of the original ones or out of a set of principal components) to represent the multidimensional space. The centroid of the i-th group will have the coordinates:

$$\begin{aligned} x_{i1} &= \sum\nolimits_1 w_{i1j}\, x_{i1j}, \qquad j = 1,\dots,N_i \\ &\vdots \\ x_{ik} &= \sum\nolimits_1 w_{ikj}\, x_{ikj}, \qquad j = 1,\dots,N_j \end{aligned} \tag{1}$$

with: $i = 1,2,\dots,n$ (n=number of clusters), N_i = number of objects in the i-th cluster, k = number of variables chosen from the data, x_{ikj} = the value of the variable k in object j of the i-th group, $w_{i1}, w_{i2},\dots, w_{ik}$ = weights used for computation of centroids.

The weight may be uniform, calculated using one variable at a time among all the variables in the data according to the formula:

$$w_{ikj} = x_{ikj} / \sum\nolimits_1 x_{ikj} \qquad j = 1 \dots,N_i \tag{2}$$

or may be given by using values of an external variable.

The standard deviation of each variable is computed according to:

$$\begin{aligned} S_{i1} &= [\sum\nolimits_1 w_{i1j}(x_{i1j} - A_{i1})^2]^{1/2} \qquad j = 1,\dots,N_i \\ S_{ik} &= [\sum\nolimits_1 w_{ikj}(x_{ikj} - A_{ik})^2]^{1/2}, \qquad j = 1,\dots,N_i \end{aligned} \tag{3}$$

where $A_{i1}, \dots, A_{ik}$ are the averages of the variables in the clusters. Euclidean distance is calculated between the centroids:

$$d_{ih} = [\sum\nolimits_1 (x_{il} - x_{hl})^2]^{1/2} \; l = 1,\dots, k \tag{4}$$

and the Student's test performed according to:

$$t_{ih} = d_{ih}/s_{ih}\, [(N_i + N_h)/N_i N_h]^{1/2} \tag{5}$$

where:

$$S_{ih} = [((N_i-1)\, S_i^2 + (N_h-1)\, S_h^2)/(N_i + N_h - 2)]^{1/2} \qquad (6)$$

with:

$$S_i^2 = \frac{1}{k}\sum_l S_{il}^2 \quad S_h^2 = \frac{1}{k}\sum_l S_{hl}^2 \quad l=1,...,k$$

These are mean squared estimates of dispersion in the data, around the two centroids. S_{il}, S_{hl} are computed by formula (3). (N_i+N_h-2) are degrees of freedom.

The probability of an object belonging to a cluster i is given by:

$$P(\mathbf{x}_{ij}) = \frac{1}{[\det(V)]^{1/2}\,(2pg)^{k/2}}\, e^{-.5[(x_{ij}-x_{ija})'\,(V)^{-1}(x_{ij}-x_{ija}))]} \qquad (7)$$

where $\mathbf{x}_{ij}$ is the vector describing the object j in cluster i and $\mathbf{x}_{ija}$ is the centroid of the cluster i, $\mathbf{V}$ is the variance covariance matrix and pg= 3.1415...

Since the domain of this function is infinite, the probability that an element belongs to a cluster is computed by integrating the function till a boundary value corresponding to a fixed probability and by checking if the element is outside or within the boundary.

If the k variables are orthogonal the probability is computed by:

$$P'(\mathbf{x}_{ij}) = \frac{1}{\prod_{l=1}^{k} S'_{ijl}\,(2pg)^{k/2}}\, e^{-.5\left[\sum_{l=1}^{k}\left(\frac{x'_{ijl}-x'_{ijla}}{S'_{ijl}}\right)^2\right]} \qquad (8)$$

In this case the integration is easier and the semiaxes are exactly $z(H_0)$ times the dispersion s' calculated for each variable.

This because $\frac{x'_{ijl}-x'_{ijla}}{s'_{ijl}}$ is the variable z.

To compute the semiaxes of the ellipsoids, three options are available:

1) The semiaxes are determined by integrating the multinormal distribution from 0 to the z-value corresponding to the selected H_0 probability. To obtain the actual semiaxes, this value is multiplied by the standard deviation of the selected variables.
2) The z value is given in input by the user; in this case the program computes the associated P_0. As a consequence the measure of the compactness associated with a z-value of each cluster is obtained.
3) The semiaxes are assumed as the maximum absolute distances of points from the centroid.

The hypervolume is conditioned by the choice of the option. The formula used for computing the hypervolume is the following:

$$HV = \prod_l \varepsilon_l\, \pi^{n/2}/\Gamma(1+n/2) \quad l=1,...,k$$

where ε_l is the l-th semiaxis of the ellipsoid and Γ is the gamma function. For each cluster TANTIT gives also the relative hypervolume as the ratio between its hypervolume and the hypervolume of the largest cluster.

The overlap between two ellipsoids is determined by considering the number of elements of one cluster included in the ellipsoid of the other one. The formula for overlap is:

$$O_{ih} = (n_i + n_h)/(N_i + N_h)$$

where n_i and n_h are respectively the number of points of cluster i included in the ellipsoid h and the number of points of the cluster h included in the ellipsoid i, N_i and N_h are the number of elements in the two clusters.

TANTIT computes two measures of overlap: the actual and the simulated one. This is obtained by using Monte Carlo simulation to distribute randomly 1000 points in each of the theoretical ellipsoid corresponding to the clusters. TANTIT produces minimum spanning trees and the dendrogram tables on the basis of distances between centroids, Student's t and probabilities of Student's t. The distances are transformed in similarity by the formula $(D_{max}-D_{ih})/D_{max}$. The ellipsoids are defined at the probability level that leaves outside a given percentage of points.TANTIT indicates the number and the percentage of points outside and inside the ellipsoids of equal concentration.The difference between the percentage of points actually outside, and the percentage of points that in theory should be outside may be used as a rough estimate of the non-linearity affecting each cluster. The level that leaves outside 5% of the points has been chosen for computing the parameters of the ellipsoids (option 1).

Applications

Two sets of data have been used to illustrate the performance of TANTIT, one (Table 1) is obtained from a simulated coenocline by program SPAGHET (Lagonegro 1984) with noise level set to 30% level. The simulated data matrix represents an ecological space as those used in indirect gradient analysis (Whittaker 1967).The matrix in Table 1 is rearranged according to 5 clusters obtained by cluster analysis by average linkage within the merged groups based on Euclidean distance. The other data set (Table 2) includes 51 relevés of a pasture on the Karst in Gorizia Province. The relevés are described by the average indicator values of Landolt (1977). Each relevé has been done within 1 m^2 enclosure. The data collection, which is part of a broad project aimed to analyze the effects of mineral fertilization on vegetation. structure and production, followed a systematic sampling design . The species cover was estimated by visual inspection. Five clusters of relevés have been defined by the same clustering method used for the coenocline (Table 1).

FIVEPA (Feoli, Lagonegro and Orlóci 1991) has been applied to define the clusters,TANTIT for testing the fit of clusters with ellipsoids, their separation and to measure their hypervolume. Principal component

Table 1. Data set of a simulated coenocline structured according to the results of cluster analysis (see text). Codes: c=clusters, p=position of the relevés in the coenocline, d=data, x=relevé left outside the corresponding ellipsoid.

c	p	d																					x
	34	2	8	5	0	8	1	0	7	6	8	0	1	2	4	0	1	3	3	0	4	2	
	37	2	8	5	1	8	1	0	7	6	8	0	1	2	4	0	1	3	3	0	4	2	
	36	2	8	5	0	7	1	0	8	6	8	0	1	2	4	0	1	3	3	0	4	2	
	40	2	8	5	1	8	1	0	8	5	8	0	1	2	4	0	1	3	3	0	4	1	
	35	2	8	5	0	7	1	0	7	6	7	0	1	2	4	0	1	2	3	1	4	2	
	38	2	7	4	1	8	1	0	7	5	7	0	1	2	4	0	1	3	3	0	4	1	x
	31	2	8	4	0	7	1	0	6	6	8	0	1	2	4	0	1	3	3	1	3	2	
	39	2	8	5	0	8	1	1	6	6	8	0	1	2	4	0	1	3	3	0	3	2	
1	33	2	7	4	1	7	1	1	6	7	8	0	1	2	4	0	1	3	3	0	4	2	x
	30	2	8	5	1	8	1	0	7	6	8	0	1	2	4	0	1	2	2	1	4	2	
	32	2	7	5	1	9	1	0	7	6	8	0	1	2	4	0	1	2	2	1	4	2	
	28	2	8	5	1	7	2	0	7	6	8	0	1	2	4	0	1	2	3	1	3	2	
	29	2	8	5	1	7	1	0	7	6	8	0	1	2	4	0	1	2	2	0	3	2	
	27	2	8	5	1	8	2	0	7	5	8	0	1	1	4	0	1	2	3	1	3	2	
	26	2	7	5	1	7	1	0	8	4	7	0	1	1	4	0	1	2	2	1	4	2	
	21	2	8	5	1	6	2	0	7	5	8	0	1	1	3	1	1	2	2	1	3	2	x
	24	2	8	5	1	7	2	0	8	5	8	0	1	2	3	1	1	2	2	1	3	1	x
	22	2	7	5	1	6	1	0	7	5	8	0	1	1	4	1	1	2	2	1	3	1	x
	12	2	8	5	2	6	1	0	5	5	8	0	1	1	3	1	1	1	2	1	2	1	
	16	2	8	5	2	6	2	0	5	4	8	0	1	1	4	1	1	1	2	1	2	1	x
	13	2	8	5	2	6	1	0	5	5	7	0	1	2	4	1	1	1	2	1	2	1	
	15	2	8	4	2	7	1	0	5	5	7	0	1	1	3	1	1	1	2	1	2	1	
	14	2	8	4	2	6	1	0	4	5	7	0	1	1	4	1	1	1	2	1	3	1	
2	11	2	8	5	2	7	1	0	5	3	7	0	1	1	3	1	1	1	2	1	3	1	
	17	2	8	4	1	7	1	0	5	5	8	0	1	1	3	1	1	2	2	1	3	1	
	20	2	8	5	1	7	1	0	5	4	8	0	1	1	4	1	1	2	2	1	3	1	
	19	2	7	5	1	7	2	0	5	5	8	0	1	1	3	1	1	2	2	1	3	1	
	18	2	8	4	1	6	2	0	5	5	8	0	1	1	4	1	1	2	2	1	3	1	
	25	2	8	4	1	6	2	0	6	4	8	0	1	1	4	1	1	2	3	1	3	2	x
	23	2	7	4	1	7	2	0	6	5	7	0	1	1	4	1	1	2	2	1	4	2	x
	45	2	8	2	0	5	1	0	8	6	8	0	1	3	4	0	1	2	4	0	5	1	
	46	2	8	2	0	6	1	1	8	6	8	1	1	3	4	0	1	2	4	0	5	1	
	50	2	6	2	0	6	1	1	8	7	8	1	1	3	4	0	1	2	4	0	5	1	x
	44	2	8	3	0	7	1	1	7	7	8	0	1	3	4	0	1	2	3	0	5	1	
3	47	2	7	2	0	5	1	0	6	5	8	1	1	3	4	0	1	3	4	0	5	1	
	49	2	7	2	0	6	1	1	6	5	8	1	1	3	4	0	1	2	4	0	5	1	
	48	2	7	2	0	6	1	0	7	6	8	1	1	3	4	0	1	3	4	0	4	1	
	43	2	7	2	1	6	1	1	7	5	8	0	1	2	4	0	1	3	3	0	4	1	
	41	2	8	2	1	7	1	0	8	6	8	0	1	2	4	0	1	3	4	0	4	1	
	42	2	7	2	1	8	1	1	8	5	7	0	1	2	4	0	1	3	3	0	5	1	x
	1	2	7	3	2	2	1	0	1	3	6	0	1	1	3	1	0	1	1	1	3	1	
4	4	2	7	3	2	2	1	0	2	3	6	0	1	1	3	1	0	1	1	1	3	1	
	2	2	7	3	2	2	0	0	2	3	6	0	1	1	2	1	0	0	1	2	3	1	x
	3	2	8	3	2	3	1	0	2	3	7	0	1	1	2	1	0	1	1	1	3	2	
	6	2	8	3	2	2	1	0	5	3	7	0	1	1	3	1	1	1	2	1	2	1	
	10	2	8	4	2	2	1	0	4	3	7	0	1	1	4	1	1	1	2	1	2	1	
5	8	2	8	4	2	3	1	0	5	3	8	0	1	1	3	1	1	1	2	1	2	1	x
	7	2	8	3	2	3	1	0	5	3	7	0	1	1	4	1	1	1	1	1	2	1	
	5	2	8	3	2	3	1	0	4	3	8	0	1	1	3	1	1	1	1	1	3	1	
	9	2	7	5	2	3	1	0	4	3	8	0	1	1	3	1	1	1	2	1	4	1	x

Table 2. Description of 51 relevés of a grassland in the province of Gorizia (NE Italy) by the Landolt's average ecological indicator values. The relevés are grouped by clusters obtained by cluster analysis (see text). The last column of the table indicates the number of the cluster to which each relevé belongs.

1.94	3.46	2.52	2.93	3.72	3.81	3.93	3.13	1
1.98	3.47	2.49	2.98	3.82	3.89	3.93	3.06	1
2.04	3.52	2.57	2.87	3.72	3.71	4.02	3.11	1
1.85	3.61	2.52	2.94	3.66	3.71	3.91	3.13	1
1.85	3.50	2.48	2.79	3.51	3.77	4.07	3.09	1
1.87	3.58	2.40	2.76	3.69	3.83	3.84	3.00	1
1.96	3.59	2.30	2.76	3.73	3.79	3.86	2.88	1
2.07	3.68	2.42	2.82	3.61	3.84	3.91	2.93	1
1.97	3.53	2.38	2.90	3.71	3.83	3.68	2.97	1
1.94	3.71	2.52	2.74	3.85	3.79	3.98	2.94	1
2.16	3.46	2.50	2.89	3.84	3.78	3.84	2.94	1
2.10	3.39	2.39	3.00	3.86	3.83	3.77	3.03	1
2.20	3.45	2.67	2.98	3.95	3.67	3.71	3.00	1
2.29	3.59	2.69	3.02	3.93	3.67	3.91	2.95	1
2.03	3.57	2.62	2.97	3.97	3.70	3.82	3.14	1
2.15	3.22	2.23	3.06	3.81	3.67	3.56	3.01	1
2.29	3.76	2.41	2.94	3.86	3.74	4.05	2.91	1
2.16	3.57	2.64	3.04	3.84	3.86	4.07	2.92	1
2.35	3.67	2.57	2.96	3.89	3.50	4.19	3.11	1
2.06	3.85	2.44	3.07	3.89	3.74	3.85	3.13	1
2.16	3.74	2.46	3.02	3.98	3.75	3.87	3.06	1
2.01	3.81	2.24	3.03	3.87	3.95	3.95	3.11	1
2.05	3.84	2.21	2.96	3.74	3.68	4.08	3.11	1
1.97	3.18	2.56	2.52	3.49	3.43	3.51	3.09	1
1.90	2.84	2.73	2.81	3.60	3.97	3.90	2.97	1
1.78	3.70	2.09	2.78	3.70	3.95	3.80	3.09	2
1.83	3.53	2.20	2.81	3.80	3.95	3.99	3.05	2
1.82	3.48	2.19	2.81	3.94	3.94	3.72	3.04	2
1.71	3.53	2.03	2.69	3.56	3.98	3.79	3.00	2
1.78	3.54	2.16	2.60	3.65	3.94	3.71	2.85	2
1.63	3.58	2.09	2.66	3.37	3.76	3.71	2.76	2
1.53	3.52	2.06	2.52	3.47	3.97	3.54	3.13	2
1.62	3.76	2.10	2.87	3.50	3.96	4.13	3.21	3
1.78	3.71	2.27	2.77	3.55	3.88	4.15	3.14	3
1.42	3.69	2.08	2.55	3.31	3.93	4.04	3.23	3
1.83	3.30	2.66	2.79	3.18	3.64	4.55	3.24	3
1.64	3.71	2.66	2.72	3.08	4.00	4.10	2.96	3
2.72	3.34	3.28	3.06	4.21	3.54	3.84	2.99	4
2.57	3.15	3.41	3.00	4.07	3.77	3.61	3.00	4
2.30	2.90	3.47	2.74	4.29	3.98	3.64	2.98	4
2.34	2.95	3.10	2.81	3.95	3.62	3.96	2.99	4
2.54	3.41	2.99	3.30	4.03	3.74	3.72	3.01	5
2.41	3.30	2.99	3.12	3.77	3.76	3.71	2.99	5
2.25	3.34	2.82	3.30	3.82	3.94	3.80	2.99	5
2.44	3.26	2.68	3.11	3.99	3.90	3.58	3.06	5
2.34	3.37	2.90	2.93	3.93	3.60	3.93	3.02	5
2.31	3.47	2.85	2.93	3.98	3.68	3.83	3.01	5
2.49	3.39	3.05	3.04	4.07	3.56	3.84	3.00	5
2.50	3.48	2.86	3.03	3.95	3.58	3.95	3.01	5
2.18	3.50	2.98	3.00	3.68	3.68	3.85	2.68	5
2.24	3.30	2.73	3.00	3.59	3.68	3.97	3.01	5

analysis, the algorithm included as an option in the program, has been chosen to reduce the dimensions of the space. The four dimensional space given by the first four principal components has been used to compute the ellipsoids in both the applications. Measures of the hypervolumes of the clusters and their relative overlap have been computed by setting to 5% the number of points left outside the ellipsoids.

Results

Figure 1 displays the minimum spanning tree (MST) of the 5 clusters obtained for the relevés of the simulated coenocline. The MST is based on the distance between the centroids.The sequence of the clusters along the gradient is perfectly reproduced. On the basis of t criterion the distance between the clusters is highly significant. This suggests that they may represent 5 different communities along the simulated unidimensional gradient.The percentages of relevés included and not included in the ellipsoids, the relative hypervolumes of the clusters are given in Table 3. The overlap, simulated and real, is given in Table 4. The percentage of points that are left outside the ellipsoids is ranging between 20 and 33%. If we consider that the ellipsoids are computed by leaving outside a percentage of 5% of the points, the clusters may be considered affected by a percentage of non- linearity ranging betweeen 15 and 28.

The MST of the five clusters of Table 2 is presented in Figure 2. Table 5 presents the average indicator values for each cluster, the hypervolume of the cluster, the percentage of relevés inside and outside the ellipsoids and the species which characterize the clusters. The percentage of releves outside and inside the ellipsoids indicates that the clusters of this data set are more affected by non-linearity than those of simulated coenocline. The distances between the five clusters are all significant according to the t criterion. Table 6 shows the overlap between the clusters. Only contiguous clusters in MST are overlapping. The cluster 3 does not show overlap proving to represent a peculiar microenvironment in the grassland. The MST shows that the clusters are arranged in the space according to a clear gradient of humidity and dispersion. These results prove that the heterogeneity of the grassland is mainly due to the soil dispersion which controls the water availability and therefore the microthermic excursions. Cluster 3 is the largest, *i.e.* environmentally

Table 3. Relative hypervolume (Hv), number of relevés (n), percentage of relevés inside (%I) and percentage of releves outside (%O) of the ellipsoids corresponding to the 5 clusters of the simulated coenocline in Table 1.

	Hv	n	%I	%O
5	0.56	10	80	20
4	1.0	18	72	28
3	.34	12	75	25
2	.03	6	67	35
1	.22	4	75	25

Table 4. Matrix of relative overlap between the five clusters of the simulated coenocline, a=based on data, b=based on simulation.

	5	4	3	2	1
	3	1	2	5	4
3a	1	0	0	0	0
3b		0	0	0	0
1a		1	0.07	0	0
1b			0.08	0	0
2a			1	0	0
2b				0.04	0
5a				1	0
5b					0
4					1

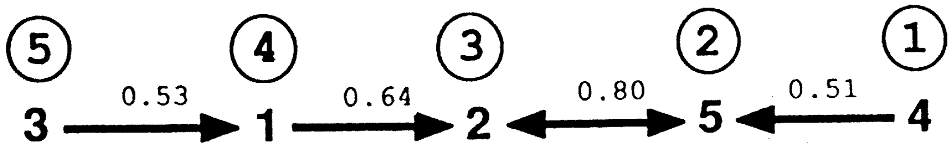

Figure 1. Minimum spanning tree of the centroids of the 5 clusters of the simulated coenocline in Table 1. The arrows indicate the direction of maximal similarity. The numbers surrounded by a circle indicate the order of the centroids along the coenocline.

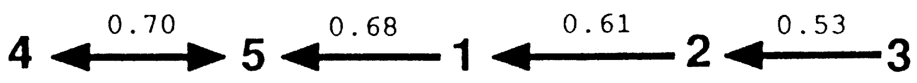

Figure 2. Minimum spanning tree of the centroids of the clusters in Table 2. The arrows indicate the direction of maximal similarity.

Table 5. Means and standard deviations of the average ecological indicator values in the 5 clusters of the grassland (see Table 2), their relative hypervolume, % of releves inside and outside the ellipsoids, and the list of species characterizing the clusters.

cluster codes	3	2	1	5	4
number of elements	5	7	25	10	4
1)Humidity	1.65 ± .16	1.73 ± .11	2.05 ± .14	2.37 ± .12	2.48 ± .19
2)pH	3.63 ± .19	3.55 ± .07	3.54 ± .22	3.38 ± .08	3.08 ± .20
3)Nutrients	2.35 ± .28	2.11 ± .07	2.47 ± .14	2.88 ± .12	3.31 ± .16
4)Humus	2.74 ± .12	2.69 ± .11	2.91 ± .13	3.07 ± .13	2.90 ± .15
5)Dispersion	3.32 ± .20	3.64 ± .19	3.78 ± .14	3.88 ± .16	4.13 ± .15
6)Light	3.88 ± .14	3.92 ± .07	3.75 ± .12	3.71 ± .13	3.72 ± .19
7)Temperature	4.19 ± .20	3.75 ± .13	3.88 ± .16	3.81 ± .12	3.76 ± .17
8)Continentality	3.15 ± .11	2.98 ± .13	3.03 ± .08	2.97 ± .10	2.99 ± .01
Hypervolume	1.00	0.15	0.78	.17	.51
% outside	20	28	20	40	75
% inside	80	72	80	60	25
Genista sylvestris	+				
Koeleria splendens	+				
Festuca rupicola		+			
Genista tinctoria		+			
Bromus erectus			+		
Brachypolium pinnatum			+		
Arrhenatherum elotius				+	
Trisetum florescens				+	
Dactylis glomerata					+
Lolium perenne					+

more heterogeneous. Cluster 4 is the most affected by non-linearity having 75% of the relevés out of the ellipsoid. The two clusters representing more specialized niches are the second and the fifth. One may be interprested as representing the niche of the community characterized by *Festuca rupicola* and *Genista tinctoria*, the other the community niche characterized by *Arrhenatherum elatius* and *Trisetum flavescens*. In this case it does not happen what usually happens in natural communities, *i.e.* that the extreme communities along the gradients are the most homogeneous (see Feoli and Lagonegro 1982).This would suggest that the Warming's statement of relative uniformity of extreme habitats (see McLean and Ivimey-Cook 1973) is not valid in the artificial environment as that of the grassland under study.

Table 6. Overlap between the clusters of Table 2, a=based on data, b=based on simulation.

	3	2	1	5	4
3a	1	0	0	0	0
3b		0	0.13	0	0
1a		1	0.01	0	0
1b			0.15	0	0
2a			0.005	0.14	0
2b			1	0.11	0
5a				1	0.14
5b					0.003
4					1

Conclusion

The application of TANTIT has shown that the use of linear models in presence of evident non-linearity is a useful exercise to investigate the structure of a data set. A simple measure of non linearity affecting different clusters is suggested. This can be used to test the efficiency of different classifications in defining clusters with linear data structures. This would be a suitable criterion for breaking up the n-dimensional space as suggested by Wildi (1989a, 1989b). Searching for linearity in n-dimensional space should be useful to identify subsystems to which linear models would be appropriately applied.

Aknowledgements. We tank prof. L. Orlóci for reading and commenting the paper and Sun Cheng Yong for useful discussions. The research was supported by Italian MPI.

REFERENCES

Anderberg M. R. 1973. *Cluster analysis for applications.* Academic Press, New York.

Digby P. G. N. and R. A. Kempton. 1987. *Multivariate analysis of ecological communities.* Chapman and Hall, London.

Feoli, E. and M. Lagonegro. 1982. Syntaxonomical analysis of beech woods of Apennines based on the program package IAHOPA. Vegetatio 50: 129-174.

Feoli E. , P. Ganis and Z. Woldu. 1988. Community niche, an effective concept to measure diversity of gradients and hyperspaces. Coenoses 3: 79-82.

Feoli E. , M. Lagonegro and L. Orlóci. 1984. *Information analysis of vegetation data*. Junk, The Hague.

Feoli E. , M. Lagonegro and L. Orlóci. 1991. FIVEPA. A computer package designed to compare sets by information measures. In E. Feoli and L. Orlóci 1991 (eds.) *Computer assisted vegetation analysis.* Kluwer, Dordrecht. pp. 429-435.

Goodall, D. W. 1964. A probabilistic similarity index. Nature 203: 1098.

Goodall, D. W. and E. Feoli. 1988. A new similarity index based on probability. Biometrics 22: 883-907.

Goodall, D. W. 1968. Affinity between an individual and a cluster in numerical taxonomy. Biom. Prax. 2: 52-55.

Goodall, D. W. and E. Feoli. 1988. Application of probabilistic methods in the analysis of phytosociological data. Coenoses 3: 1-10.

Lagonegro, M. 1984. SPAGHET: A coenocline simulator useful to calibrate software detectors. Studia Geobotanica 4: 63-99.

Lagonegro, M. and E. Feoli. 1985. The use of ellipses of equal concentration to analyze ordination vegetation pattern. Studia Geobotanica 5: 143-165.

Landolt E. 1977. Ökologische Zeigerwerte zur Schweizer Flora. Ber. Geobot. Inst. ETH, 64: 64-207.

Legendre L. and P. Legendre 1983. *Numerical Ecology.* Elsevier, Amsterdam.

McLean R. C. and W. R. Ivimey-Cook. 1973. *Textbook of theoretical botany.* Longman, London.

Mardia K. V. , J. T. Kent and J. M. Bibby 1979. *Multivariate analysis*. Academic Press, New York.

Orlóci, L. 1978. *Multivariate analysis in vegetation research.* 2nd ed. Dr. W. Junk, The Hague.

Wildi, O. 1989a. A new numerical solution to traditional phytosociological tabular classification. Vegetatio 81: 95-106.

Wildi O. 1989b. Analysis of the disintegrating group and gradient structure in Swiss riparian forests. Vegetatio 83: 179-186.

Whittaker, R. H. 1967. Gradient analysis of vegetation. Biol. Rev. 42: 207-264.

Manuscript received: March 1992

Appendix

Program TANTIT

This program is written in FORTRAN 77 for PC and VAX. The data in input (variables as the columns) may be structured according to groups or unstructured. A column vector of labels, which are in one to one correspondence with the objects in the table, is used to assign each object to one of the groups.

TANTIT allows the comparison between groups by using, optionally, one to five variables simultaneously.

The user can chose to transform the data table by PCA, the algorithm is included in the program.The new scores can be written to a file, for further use. The distance matrix between the centroids of the clusters, expressed as an option, by Euclidean distance, by t or t probability is used to find the MST and the dendrogram by single linkage.These matrices (lower triangular part) may be saved on file. The program computes the hypervolumes of the ellipsoids corresponding to the groups and the overlap between them according to real and simulated data.