COENOSES 4 (2): 95-106, 1989

FUZZY CLUSTERING OF ECOLOGICAL DATA

S. Marsili-Libelli, Department of Systems and Computer Science, University of Florence, via S. Marta 3, 50139 Florence, Italy

Abstract. Ordination and classification have always been important stages in ecological data analysis. This paper presents a clustering technique based on fuzzy sets to obtain both ordination and classification particularly well suited for ecological analyses. Three algorithms are presented to categorize data, classify new ones, and produce fuzzy dendrograms. Some examples demonstrate the algorithms' capabilities to yield graded classification of data and show that this approach has more flexibility compared to classical methods.

1. Introduction

Clustering represents a widely used technique to process data and to establish connections within seemingly unrelated observation sets. Deterministic clustering in a sweeping variety of forms has been used to process any kind of experimental data from the most diverse disciplines: for a thorough survey refer to Cormack (1971), Orlóci (1978), Pielou (1984), and Dillon and Goldstein (1984). However, when dealing with biological or ecological data which bear a large inherent uncertainty, such deterministic groupings often result in misleading or utterly unrealistic associations. Further, the hard cluster has no provision to deal with borderline situations, which are very common in ecological analysis, warning the scientist to what extent a particular cluster is contrived.

It will be shown in this paper that fuzzy clustering is by its very nature able to determine the degree of membership of a data point to a certain group, thus giving more information regarding the extent to which the given data belong to a particular structure. Fuzzy sets were first introduced by Zadeh (1965) and for a thorough survey on fuzzy sets and their use in the expert systems context the reader is referred to Gaines and Kohout (1977), Kaufmann and Gupta (1985), Negoita (1985), Zimmermann (1985), Kandel (1986), and Yager et al. (1987). In the present context of ecological data processing, the notion of fuzziness will be regarded primarily as a means to quantify the degree of similarity of a given data point with respect to previously grouped data (Feoli and Zuccarello 1986, 1988).

Many different multivariate techniques go under the general name of clustering, though there is little general agreement on what actually constitutes a cluster. In this context the notion of cluster will be used as an aggregation procedure in which a definition of similarity is used to categorize multivariate date. Clustering algorithms can be divided into two broad categories: agglomerative (or hierarchical) and divisive (or partitioning). The former aggregates the data according to a varying similarity threshold, whereas the latter partitions the data with respect to a similarity-sorting criterion. This paper will be concerned with both categories and will show how the fuzzy approach provides an unifying basis for these seemingly contrasting approaches using a sorting criterion based on fuzzy sets. The main advantage with respect to conventional clustering techniques is that this approach produces a graded membership for the classified data, thus enabling the user to discern the degree of aggregation within the given data set.

The paper is organized as follows: section 2 reviews the classical deterministic divisive clustering involving the minimization of a membership functional. Section 3 extends the previous method to graded partitions through a *fuzzy clustering* algorithm, obtained with the *fuzzification* of the previous membership functional. Then section 4 deals with fuzzy cluster validity in term of partition efficiency, discussing the selection of the optimal number of clusters. Section 5 shows how the fuzzy algorithm of section 3 provides guidelines for the construction of a *fuzzy classifier* of new data with respect to existing clusters. In section 6 the familiar dendrogram representation of hierarchical divise clusters is revisited and the new concept of *fuzzy dendrogram* is introduced. In section 7 some significant examples are presented to demonstrate that this approach is particularly suited for ecological applications.

2. A review of deterministic partitions

The two main constituents of deterministic partitioning techniques are a *similarity measure* and a *membership functional* whose minimization should produce the optimal partition. This is usually an iterative process and presently it is assumed that the number of clusters is predetermined by the user, whereas the problem of finding the most appropriate number of clusters is dealt with later in the paper. Let the data points be represented by a matrix $\mathbf{X} \in R^{N \times \ell}$, where N is the number of data and ℓ the dimension (or number of features) of each of them, *i.e.*

$$\mathbf{X} = \{\mathbf{X}_k \mid k=1, \ldots, N \mid \mathbf{X}_k \in R^{\ell}\} \quad (1)$$

Now, the data set $\mathbf{X}$ is to be partitioned into C clusters. To obtain the required similarity measure assu-

me that i = 1, 2, ..., C clusters have been formed from the original data set and that the mean of the j-th feature (j = 1, 2, ..., ℓ) of the data grouped into the i-th cluster is expressed by V_i (j). Then the *similarity* of the k-th individual to the i-th cluster center V_i (j) can be defined in terms of all the ℓ features of the k-th point $\mathbf{X}_k$. Several definitions of similarity can be adopted (Cormack 1971) to group the data. In this paper the euclidean distance between the k-th point $\mathbf{X}_k$ and the i-th cluster V_i will be assumed to represent such a measure, *i.e.*,

$$D^2_{k,\,i} = \sum_{j=1}^{\ell} [X_k\,(j) - V_i\,(j)]^2 \tag{2}$$

whereas the partition functional is defined as the sum of distances defined by (2), i.e.

$$J_d\,(C) = \sum_{i=1}^{C} \sum_{k=1}^{N} D^2_{k,\,i} \tag{3}$$

Minimizing J_d (C) amounts to minimizing the within-group squared distances, so that the data which have the least dissimilarities are allocated to the same cluster. Starting with a predetermined number of clusters C and a tentative partition, J_d (C) is minimized iteratively moving the data from one cluster to another until a minimum is reached. The mutually farthest C points are an obvious choice as the starting centroids, which are then recomputed at each iteration until a minimum of J_d (C) is obtained. At the end of the process the membership matrix $\mathbf{U} \in R^{C\times N}$ of the data points $\mathbf{X}$ with respect to the clusters will be composed of 1 or 0 entries so that the memberships of each point $\mathbf{X}_k$ to the C clusters are mutually exclusive.

3. A fuzzy partition algorithm

The approach just outlined produces a "hard" partition of the data set $\mathbf{X}$ into C predetermined clusters, so that a point $\mathbf{X}_k$ is eventually assigned to one and only one cluster. Conversely, the idea of using fuzzy sets to represent graded partitions stems from the inherently uncertain nature of ecological data and the consequent inability to obtain reliable results in deterministic terms. When classifying ecological affinities it is often impossible or at least misleading to impose hard assignments, whereas a gradual transition from a high to a low affinity is often much more appropriate. Fuzzy sets are by their very nature able to express the degree of membership of a data point to a number of classes, thus quantifying the extent to which the given data can be associated to a particular group. This is an all-important feature in ecological data processing where system complexity often adds up to data scarcity with the result that the boundaries between differing ecological situations are sometimes very difficult to extricate.

The original idea of using fuzzy sets in clustering techniques came from Ruspini (1969), to be later extended by Bezdek (1974, 1981), Dunn (1974a), Gustafson and Kessel (1979). Basically there are two ways of introducing fuzziness into a partition: one is to fuzzify the partition functional, the other is to consider the similarity matrix as a fuzzy matrix. This latter approach has been recently developed by Zhao (1987) to determine overlapping soft clusters, whereas the former still represents the mainstream of fuzzy clustering literature and it will be adopted in this paper. The difference with the deterministic approach of section 2 is that now the partition functional to be minimized is defined as

$$J_f\,(C,\,m) = \sum_{k=1}^{N} \sum_{i=1}^{C} (u_{k,\,i})^m\, D^2_{k,\,i} \tag{4}$$

where the weights $u_{k,\,i}$ are the elements of the fuzzy membership matrix U defined as

$$\mathbf{U} = \{u_{k,\,i} \mid k=1, \ldots, N \mid i = 1, \ldots, C \mid u_{k,\,i} \in (0, 1) \mid \mathbf{U} \in R^{C\times N}\} \tag{5}$$

The exponent m $\in$ [1, 0) appearing in (4) determines the incidence of fuzzy memberships $u_{k,\,i}$ on the partition and $D^2_{k,\,i}$ is the same squared euclidean distance of eq. (2). The cluster centers $\mathbf{V}_i$ are defined as the weighted fuzzy sum of the $\mathbf{X}_k$ points

$$\mathbf{V}_i = \frac{\sum_{k=1}^{N} (u_{k,i})^m\, \mathbf{X}_k}{\sum_{k=1}^{N} (u_{k,\,i})^m} \qquad i = 1, 2, \ldots, C \tag{6}$$

It can be seen that fuzziness enters into the determination of the cluster centers $\mathbf{V}_i$ since the points $\mathbf{X}_i$ are weighted with their fuzzy membership. So the data with a high membership will attract the centroid more than points with a low membership. Furthermore, since each matrix element is bounded, i.e. $u_{k,\,i} \in (0, 1)$, increasing m results in a fuzzier, less discriminating partition. The cluster centers $\mathbf{V}_i$ represent the set of *prototypes, i.e.*, the values that the typical constituent of that cluster should have, whereas the $u_{k,\,i}$ component of the fuzzy matrix $\mathbf{U}$ denotes the extent to which the point $\mathbf{X}_k$ is akin to the i-th cluster. Contrary to the minimization of the hard functional J_d (C) of section 2, where minimization was achieved by moving data points from one cluster to another, functional J_f (C, m) is minimized by assigning suitable values to the fuzzy matrix $\mathbf{U}$ of eq. (5), reminding that the sum of the memberships of each point $\mathbf{X}_k$ to all the clusters must be

unity, *i.e.*,

$$\sum_{i=1}^{C} u_{k,i} = 1 \quad \forall\, k = 1, \ldots, N \tag{7}$$

The general solution to this problem is outlined in detail in Bezdek (1981) and a simplified version is summarized here. Minimizing the composite function obtained from eqs. (4) and (7) yields

$$F(U_k, \lambda) = \sum_{i=1}^{C} (u_{k,i})^m D_{k,i}^2 - \lambda \left(\sum_{i=1}^{C} u_{k,i} - 1 \right) \tag{8}$$

where λ acts as a Lagrange multiplier. Considering a generic element of $\mathbf{U}$, $u_{k,i}$ and setting the partial derivatives to zero yields

$$\frac{\partial F}{\partial \lambda} = \sum_{i=1}^{C} u_{k,i} - 1 = 0 \tag{9}$$

$$\frac{\partial F}{\partial u_{k,i}} = \{ m (u_{k,i})^{m-1} D_{k,i}^2 - \lambda \} = 0 \tag{10}$$

Solving eq. (10) for $u_{k,i}$ yields

$$u_{k,i} = \left\{ \frac{\lambda}{m D_{k,i}^2} \right\}^{1/(m-1)} \tag{11}$$

substituting (11) into 9 and taking summation over all the clusters C yields

$$\left(\frac{\lambda}{m}\right)^{1/(m-1)} = \left\{ \frac{1}{\sum_{i=1}^{C} \frac{1}{D_{k,i}^2}} \right\}^{1/(m-1)} \tag{12}$$

Eliminating the term $\left(\frac{\lambda}{m}\right)^{1/(m-1)}$ between eq. (11) and eq. (12) yields the fuzzy matrix $\mathbf{U}$ in which the element $u_{k,i}$ is a function of the ratio between the distance $D_{k,i}^2$, between X_k and the i-th centroid $\mathbf{V}_i$, and $D_{k,i}^2$, between the same $\mathbf{X}_i$ and all other centroids $\mathbf{V}_j$.

$$u_{k,i} = \frac{1}{\sum_{j=1}^{C} \left[\frac{D_{k,i}}{D_{k,j}} \right]^{2/(m-1)}} \tag{13}$$

The algorithm must be organized recursively as follows:

Algorithm 1:

(i) a tentative set of cluster centers $\{\mathbf{V}^0\} = \{\mathbf{V}_i^0 | i = 1, \ldots, C\}$ is determined from the C mutually farthest points according to eq. (2);
(ii) an initial fuzzy partition $\mathbf{U}^0$ satisfying the consistency constraint (7) is determined from $\{\mathbf{V}^0\}$;
(iii) a new fuzzy partition $\mathbf{U}^t$ is determined using (13);
(iv) the cluster centers $\{\mathbf{V}^t\}$ are recomputed through eq. (6) on the basis of $\mathbf{U}^t$;
(v) when two successive fuzzy partitions $\mathbf{U}^{t-1}$ and $\mathbf{U}^t$ differ by less than a given amount the iteration is terminated and $\mathbf{U}^t$ is taken as the partition which minimizes (4). In practice the termination criterion considers the maximum difference computed for each point and each cluster center

$$\delta = \underset{k}{\text{MAX}}\, \underset{i}{\text{MAX}} |u_{k,i}^{t-1} - u_{k,i}^{t}| \tag{14}$$

The problem of singularity in (13) remains to be considered. In the event that a point coincides with a centroid, then $D_{k,i}^2 = 0$ for that point. In this case eq. (13) is skipped and the corresponding $u_{k,i}$ is assigned full membership, *i.e.* $u_{k,i} = 1$.

4. Fuzzy partition efficiency

The fuzzy partition functional $J_f(C, m)$ has two algorithmic parameters: C and m. As already stated, the exponent m is a way to introduce more or less fuzziness into the partition, i.e. the higher m, the fuzzier the partition. Now we turn our attention to C, the number of clusters, to determine whether there exists for a given data set $\mathbf{X}$ a particular C which produces the most efficient partition. In fuzzy clustering two opposite requirements must coexist. Though the final goal is to produce a discriminating partition enabling unambiguous feature extraction, on the other hand fuzziness is intentionally introduced to rank data memberships. In general, though fuzziness is introduced to enhance clustering flexibility, both C and m should be such that the resulting partition is significant from an information-theoretic point of view. A widely used indicator of the information content of a cluster partition is the partition entropy (Dunn 1974b; Bezdek 1981) defined as

$$H(U, C) = -\frac{1}{N} \sum_{k=1}^{N} \sum_{i=1}^{C} u_{k,i} \log u_{k,i} \tag{15}$$

where it is agreed that $u_{k,i} \log (u_{k,i}) = 0$ when $u_{k,i} = 0$. Further, it has been shown (Dunn, 1974b) that the normalized partition entropy defined as

$$H_n(U, C) = H(U, C)/\left(1 - \frac{C}{N}\right) \tag{16}$$

is a good indicator of partition efficiency. In fact, it was suggested that the normalized partition entropy mini-

mization corresponds to a maximization of the likelihood that a given data set X may indeed contain C subsets with homogeneous features. Therefore, the indicator (16) will be used throughout to determine the optimal number of prototypes whenever their number is not defined a-priori.

5. A fuzzy classifier

It may be convenient to use the knowledge already gained in partitioning a given set to classify more data at a later time. This knowledge is in fact condensed into the prototypes $\{\mathbf{V}_i | i=1, \ldots, C\}$ and can be used to classify subsequent data without having to reprocess the composite set. This can be done in a computationally economical way if a *classifier* is developed which retains the information gained in the previous partition.

The problem is then the design of a classifier to assign fuzzy memberships to a new entry $\mathbf{X}_{N+1}$ on the basis of the prior knowledge $\{\mathbf{V}_i | i=1, \ldots, C\}$. Several approaches have been proposed at this regard. Restricting ourselves to non-Bayesian solution, Bezdek (1981) reported a nearest-prototype classifier. In that case the new point is unambiguously assigned to the cluster whose center is nearest to $\mathbf{X}_{N+1}$ according to the predefined norm (2). Thus, no fuzzy decision is involved and a hard partition results, irrespective of how the prototypes were generated. A different solution is proposed here to classify one single new entry $\mathbf{X}_{N+1}$. The fuzzy membership computation (13) can be used again to determine the memberships of the new point, after the distances between $\mathbf{X}_{N+1}$ and each of the prototypes $\mathbf{V}_i$, $\{D^2_{N+1,\,i} | i=1, \ldots, C\}$, are determined. Contrary to the iterative solution of section 3, now the prototypes $\mathbf{V}_i$ remain unchanged whereas eq. (13) is used again, although in a slightly different from, to obtain the memberships of the new point

Algorithm 2

$$u_{N+1,\,i} = \frac{1}{\sum_{j=1}^{C} \left[\frac{D_{N+1,\,i}}{D_{N+1,\,j}} \right]^{2/(m-1)}} \qquad (17)$$

Notice that condition (7), which now can be written as

$$\sum_{i=1}^{C} u_{N+1,\,i} = 1 \qquad (18)$$

implicitly holds because of the way in which (13), and therefore (17), were derived. Thus eq. (17) is the required classifier yielding the fuzzy memberships $\{u_{N+1,\,i} | i=1, \ldots, C\}$ of the new entry $\mathbf{X}_{N+1}$ with respect to the set of predetermined prototypes $\{\mathbf{V}_i | i=1, \ldots, C\}$. To decide whether the new point $\mathbf{X}_{N+1}$ is worth including in the existing basis of knowledge the partition entropy (15) still provides a guideline. If the new partition determined from the extended basis $\{\mathbf{X} \cup \mathbf{X}_{N+1}\}$ yields a lower value for H_n (U, C) then the new point $\mathbf{X}_{N+1}$ is considered enough representative to be included in the basis, otherwise the prototypes $\mathbf{V}_i$ are left unchanged.

6. Fuzzy dendrograms

When ecological data are processed through hierarchical clusters, they are grouped according to a decreasing similarity threshold and the results are often presented as dendrograms. Such agglomerative techniques are best suited for data which do not exhibit any clear substructure. In this case partition clusters, as described in sections 2 and 3, would be inappropriate and dendrograms represent a better way to rank data associations. Though in section 2 it was stated that hierarchical and divisive techniques may be viewed as opposite approaches, Algorithm 1 of section 3 can be used to produce *fuzzy dendrograms* which can be interpreted much in the same way as deterministic dendrograms produced with classical techniques.

When no underlying structure is apparent in the data set, i.e. when H_n (C) has its minimum for C=2, then Algorithm 1 can be applied to the data with C=2 and the resulting partition can be organized into a dendrogram with the following procedure:

Algorithm 3:

(1) Algorithm 1 is applied to the data set to produce two fuzzy sets $\mathbf{U}_1$, $\mathbf{U}_2$, each contaning the memberships of the N data to the two clusters.

(2) The elements of $\mathbf{U}_1$, $\mathbf{U}_2$ are sorted in terms of decreasing membership and only the element $u_{i,\,k} > 0.5$ are retained. Thus the reduced sets $\mathbf{U}_1^*$ and $\mathbf{U}_2^*$ are obtained.

$$\mathbf{U}_1^* = \{u_{1,\,1} > u_{2,\,1} > u_{3,\,1} > \ldots > u_{N_1,\,1}\} \qquad (19)$$

$$\mathbf{U}_1^* = \{u_{1,\,2} > u_{2,\,2} > u_{3,\,2} > \ldots > u_{N_2,\,2}\} \qquad (20)$$

with N_1, $N_2 < N$ and $u_{i,\,k} > 0.5$ $\forall i, k$.

The labelling order of (19) and (20) refers to the sorted sets and not to the original point labels.

(3) Points are then aggregated within each cluster $\mathbf{U}_1^*$ and $\mathbf{U}_2^*$ in terms of their membership, i.e.

$$\alpha_1 (i, j) = \text{MAX}\,(u_{1,\,i}, u_{1,\,j}) \quad (i, j = 1, \ldots, N_1) \qquad (21)$$

$$\alpha_2 (i, j) = \text{MAX}\,(u_{2,\,i}, u_{2,\,j}) \quad (i, j = 1, \ldots, N_2) \qquad (22)$$

therefore the dendrogram represents the level of decreasing similarities within each cluster.

(4) The two clusters are joined at the least significant

Table 1. Prototypes for Fisher's Iris data (Fisher 1936).

	Sepal Length	Sepal Width	Petal Length	Petal Width
I. Setosa	5.0608	3.5904	1.4906	0.3103
I. Versicolor	5.8759	2.6263	4.1672	1.2709
I. Virginica	6.8024	2.9866	5.6776	2.1436

level, using the point with the least membership in both clusters, thus completing the dendrogram.

The resulting graph connects couples of points in order of decreasing similarity. Points with a high affinity are aggregated first and more points are later joined as the similarity threshold is progressively lowered. In this context membership fuzziness plays a role similar to any distance measure in classical dendrograms. However, Algorithm 3 cannot yield non-monotonic dendrograms, as may happen with centroid-based dendrograms (Pielou, 1984) where centroid distances are used to aggregate data. In the present case, instead, memberships are used as the clustering criterion and being sorted in decreasing order no threshold reversal is possible.

7. Ecological applications of fuzzy algorithms

The previous fuzzy algorithms are now applied to ecological data sets to illustrate their usage and behaviour. The examples are organized to follow the exposure covered in the previous sections.

7.1. Fisher's Iris data

As a first application, consider the well-known Fisher's Iris data (Fisher, 1936) obtained by measuring the length and width of petals and sepals of three Iris species (I. Setosa, I. Versicolor, I. Virginica). This data set, already used by Bezdek (1981) in connection with fuzzy sets, will be used to demonstrate both fuzzy partition and classification algorithms (Algorithms 1 and 2). A subset of 10 samples for each species out of the original set (encompassing 50 data for each species) were sampled at random and processed with Algorithm 1 with $C=3$ and $m=1.5$. The resulting fuzzy partition was fully consistent with Fisher's species discrimination. In fact all the data were placed in the correct cluster with a membership greater than 0.9. The clusters appeared to be well separated and the normalized partition entropy was indeed minimal for $C=3$, thus indicating the biological consistency of the choice. Now, to demonstrate the classifier of section 5, new samples extracted from the original Fisher's set are classified with respect to the prototypes resulting from the previous partition and listed in Table 1. Three data, one for each species were selected at random from Fisher's original data set and classified with the same exponent $m=1.5$, obtaining the results of Table 2. It can be seen that all the new entries are correctly recognized and placed into the proper species set. Thus the information contained in the prototypes is sufficient for correct classification of subsequent homogeneous data.

7.2. Crop growth model parameters

A simplified crop growth model previously developed by the author is now considered. It consists of a simple growth equation describing the evolution of crop dry matter over time

$$W(t+h) = \frac{L\,W(t)\,e^{-St}}{1 + \frac{L-1}{K}\,W(t)\,e^{-St}} \qquad (23)$$

where W (t) is the total dry matter per unit area at time t and h is the time interval between successive harvests. This model, whose derivation and properties are described elsewhere (Marsili-Libelli 1985a) has three parameters {L, S, K} to adjust to various growth conditions. In particular L is the growth rate, S is a stress factor taking into account the aging mechanism which impairs the reproductive capacity of the plant as its development progresses, and K represents the carrying capacity of the environment. These parameters can be numerically estimated from crop data (Marili-Libelli 1985b). This section will pursue further the preliminary results of Marsili-Libelli (1986) in assessing the possibility that the parameters may be clustered by species and/or cultivars thus enhancing their biological significance. Moreover, since the carrying capacity K was not found to be an essential clustering feature, in order to simplify the analysis only L and S are considered as cluster parameters. The model parameters of Table 3 refer to two differing species, Barley and Soybean, with the further complication that one Soybean crop was intentionally defoliated to study the effect of a defoliating pest. This pathological condition reflects in the parameter values, particularly S which has an

Table 2. Fisher's Iris classification of new entries.

	Sepal Length	Sepal Width	Petal Length	Petal Width	Memberships Setosa	Versicolor	Virginica
I. Setosa	5.1	3.8	1.9	0.4	0.9991	0.0001	0.0008
I. Versicolor	5.0	2.0	3.5	1.0	0.0246	0.9230	0.0524
I. Virginica	6.9	3.1	5.1	2.3	0.0004	0.9857	0.0139

Table 3. Growth model parameters from Barley and Soybean crop data (Marsili-Libelli 1985b).

n	Symbol	Species	L	$S \times 10^3$ (d^{-1})
1	(A)	Barley c.v. Onice[a]	1.0615	0.8606
2	(A)	Barley c.v. Onice[a]	1.0656	0.8119
3	(A)	Barley c.v. Robur[a]	1.0620	1.2344
4	(A)	Barley c.v. Robur[a]	1.0500	0.6011
5	(B)	Soybean[b]	1.1031	0.4742
6	(B)	Soybean[c]	1.0975	0.4582
7	(B)	Soybean[c]	1.1265	0.7080
8	(C)	Soybean[b]	1.0691	0.2026

Source: a) M. Odoardi, Ist. Sperimentale per la Cerealicoltura, Fiorenzuola (Italy), personal communication, 1984. b) Florida (Quincy) experimental crop, (Ingram et al. 1981). c) Florida (Gainsville) experimental crop, (Wilkerson et al. 1983).

abnormally low value. Several numerical experiments are now carried out on this data set which includes normal as well as pathological data.

7.2.a Parameters clustering

First, a comparison is performed between the fuzzy clustering procedure of section 3 and some well known deterministic partitions, based on different distance measuremetns. For these, the implementation provided by STATGRAPHICS, a commercial statistical package, were used. These procedures work in a hierarchical fashion: given N points, N-1 clusters are formed and then they are successively merged until the prescribed number of clusters is obtained. The aggregation depends on the particular distance criterion being adopted. The following were selected: average, median, centroid, furthest neighbor, nearest neighbor, and seeded. The first five of these are the well-known hierarchical criteria described for example in Pielou (1984). The last one is a non-hierarchical deterministic method which seeks to partition the data into a predetermined number of clusters starting with an initial tentative partition (the "seed"). Thus it closely resembles the fuzzy partition algorithm of section 3, except for the lack of fuzzification.

The L and S value obtained from these crops data were processed with all these six methods and the results are shown in Fig. 1 where the symbol A refers to Barley parameters, B to Soybean and C to defoliated Soybean. Obviously two clusters were selected. With deterministic procedures, it can be seen that the first four methods (average, median, centroid, further neigh-

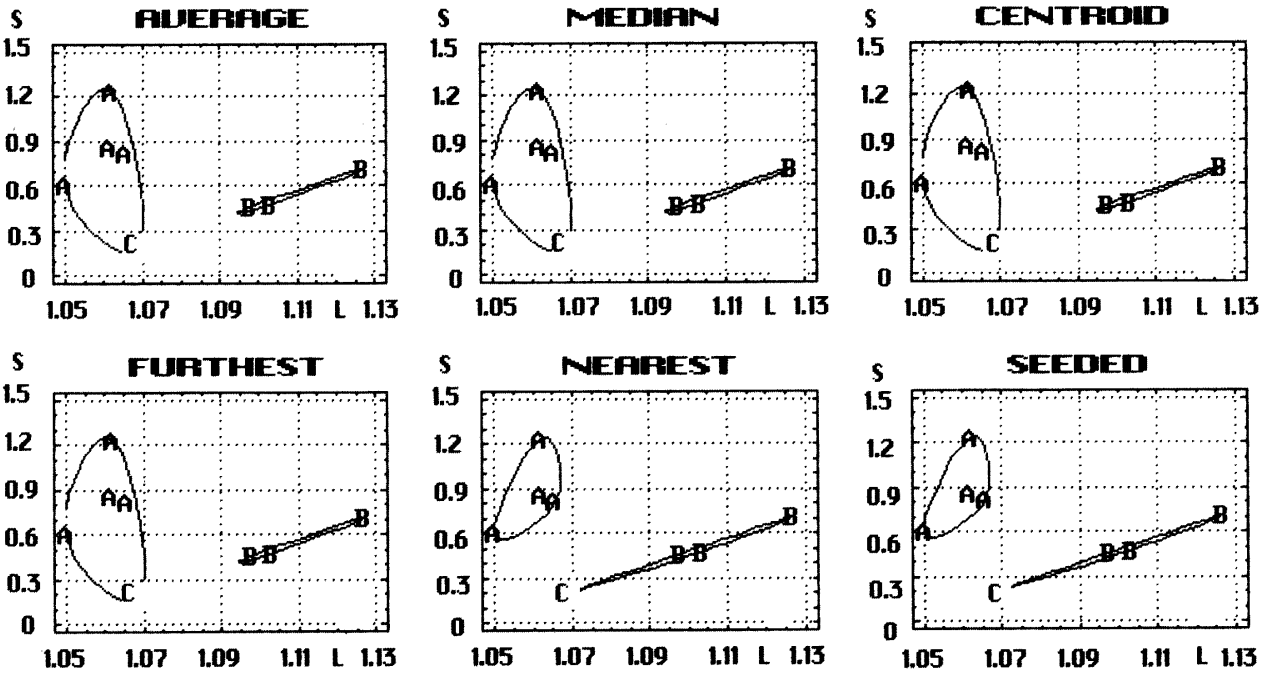

Fig. 1. Comparison of deterministic and fuzzy clustering of Soybean/Barley growth parameters. The five deterministic partitions were obtained through STATGRAFICS. (A = Barley, B = Soybean, C = Defoliated Soybean).

Table 4. Invertebrates counts in three riverbed communities (Orloci 1982).

Species	Moss beds					Cress beds					Sandy beds				
quadrat #:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Tubellaria	2	5	4	3	2	0	0	0	3	3	0	0	1	0	2
Oligochaeta	1	1	0	1	1	27	25	19	23	25	26	32	30	35	29
Isopoda	68	44	61	53	50	8	28	10	15	9	5	1	5	3	4
Amphipoda	24	21	20	25	22	25	21	23	21	20	9	7	7	6	8

bor) associate the defoliated Soybean with Barley and only the nearest neighbor and the seeded methods produce the correct clusters. In the latter case, though, the result bears no indication that one Soybean data comes from a pathological crop. Instead the fuzzy cluster is capable of such an indication through its graded membership. In fact processing the same data with the fuzzy algorithm (with m = 1.5) yields the results of Fig. 2, in which the 0.9 and 0.5 fuzziness contours are shown. The data falling within the 0.9 contour can be regarded as surely belonging to that cluster, whereas the 0.5 boundary (obviously a straight line in a 2-cluster, 2-dimensional problem) should be regarded as the maximum uncertainty locus and is in fact the discriminating line between the two clusters. The defoliated Soybean entry is clustered in a revealing way: though being placed on the correct side of the boundary, it is assigned a relatively lower membership value to signify that, though being recognized as a Soybean data, its value is not typical of that species, thus acknowledging its pathological nature.

7.2.b Prototype extraction.

Excluding the pathological Soybean parameters from the data of Table 3, Algorithm 1 is again applied to obtain typical Barley and Soybean values to be used in subsequent classification of unrelated data. Two well-separated clusters were then obtained with the following prototypes V_B = {L = 1.0598 S = .7823} and V_S = {L = 1.0975 S = .5765}, where the subscripts S and B stand for Soybean and Barley respectively. Now this information is used for further classification, entering the following data obtained from other sources: a normal Barley crop (Kershaw, 1973) and the previous defoliated Soybean crop (entry n. 8 of Table 3). New barley and soybean entries for classification:

Barley: L = 1.0559 S = 0.6158
Soybean: L = 1.0691 S = 0.2026

Using the classifier (19) on these data with the given prototypes yields the following memberships

Barley: U_B = 0.9298 U_S = 0.0702
Soybean: U_B = 0.2367 U_S = 0.7633

Notice that in this case the membership of the defoliated Soybean is slightly lower than in the previous case (U_S = 0.774), when it contributed to the determination of the centroids. From these results, it can be concluded that the new Barley data {L, S} are grouped within the Barley cluster with a very high degree of membership, *i.e.* the algorithm recognizes these parameters as surely coming from a Barley crop. Conversely, the defoliated Soybean is recognized as resembling the known Soybean knowledge, but with a much lower degree of confidence. The algorithm recognizes this new set of parameters as fairly distant from the normal Soybean data, but still loser to Soybean than to Barley.

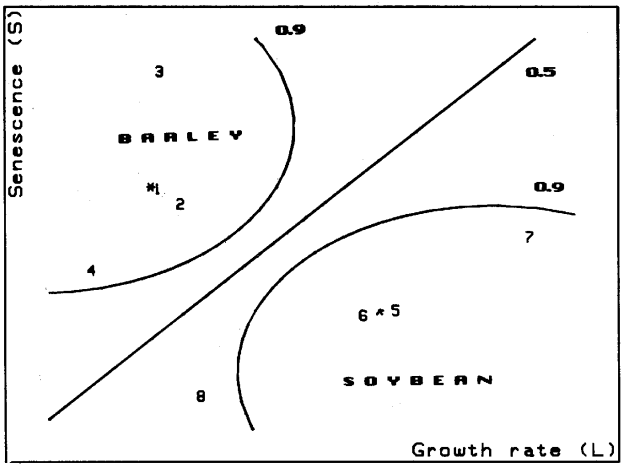

Fig. 2. Fuzzy clustering of crop growth model (22) parameters: growth rate (L) vs. senescence (S). The data from Table 3 with m = 1.5 are used and the loci of 0.9 and 0.5 membership are shown.

7.3 Clustering of riverbed communities

The distribution of several invertebrates species in three different stream bed habitats (Orlóci 1982) is now considered. In this case the data consist of species counts in three different acquatic environments. Five replicates for each habitat are reported in Table 4. To generate the initial basis of knowledge all the available quadrats were used with the exception of # 5, 10, 15. These will be used later to demonstrate the classifier. With this basis the ordination procedure yields three minimum-entropy, well separated clusters who-

Table 5. Prototypes of riverbed communities obtained with Algorithm 1 using data #1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14 of Table 4.

	Tubellaria	Oligochaeta	Isopoda	Amphipoda
V_M =	{3.501	0.834	56.321	22.491}
V_C =	{0.929	23.478	15.222	22.566}
V_S =	{0.258	30.726	3.549	7.300}

se centers are reported in Table 5. In this case the prototypes are defined in terms of the typical occurrence of each of the four species for each given habitat. Subsequent classification of the remaining quadrats #5, #10, and #15 yields the memberships of Table 6. It can be see that these three quadrats are correctly classified in the proper habitat on the basis of the existing knowledge summarized in Table 5. By contrast, if all the quadrats had been used to generate the three clusters, the prototypes would have been those of Table 7, where the three quadrats #5, #10, #15 would have been correctly placed anyway (compare with Table 5).

Table 6. Memberships of new riverbed data using the prototypes of Table 5.

New entry	Moss beds	Cress beds	Sandy beds
Quadrat #5 (Moss):	0.9734	0.0231	0.0036
Quadrat #10 (Cress):	0.0429	0.8830	0.0740
Quadrat # 15 (Sand):	0.0054	0.0491	0.9455

7.4 Eutrophication discrimination

Eutrophication has long being considered a major nuisance of water-bodies mainly triggered by phosphorus pollution. Generally the state of a water body with respect to eutrophication falls in one of four categories:

Table 7. Prototypes obtained with the complete data set.

	Tubellaria	Oligochaeta	Isopoda	Amphipoda
V_M =	{3.191	0.856	55.166	22.405}
V_C =	{1.383	23.710	13.905	22.029}
V_S =	{0.591	30.382	3.681	7.477}

Oligotrophic, Mesotrophic, Eutrophic and Ipertrophic, according to the extent of pollution, which is assessed through three main parameters: transparency (measured with a Secchi disk), average annual phosphorus concentration (mg m^{-3}) and average annual chlorophyll-a concentration (mg m^{-3}). Obviously there is no clear-cut transition from one class to another and many intermediate situations may occur. This is therefore the typical context in which fuzzy clustering can be successfully applied. In this case the goal is to determine the fuzzy sets describing the four main eutrophication stages, and obtain the related prototypes which may be used for independent classification of more data. The data of Table 8 from OECD (Marchetti, 1987) were used, spanning the whole range of the main three watch variables. Clustering these data with C=4 and m=1.5 yields four well separated clusters, whose memberships are plotted in Fig. 3 to show that the resulting profiles are in fact pertinent to the observed transitions from one stage to the next. In fact they compare well with OECD probabilistic curves (Marcheti 1987). The determination of membership functions through Algorithm 1 can be regarded as a substitute for the empirical membership function determination describing the relation between a fuzzy set and its support (Zimmerman 1985). Furthermore, the prototypes listed in Table 9 were also obtained. These values can be considered as typical of each eutrophication stage and can be used to assess the degree of eutrophication of new samples.

Table 8. Eutrophication data.

Sample	Chlorophyll-a (mg m^{-3})	Phosphorus (mg m^{-3})	Transparency (m)
1	0.5	1.0	20
2	1.0	2.5	15
3	2.5	5.0	10
4	5.0	7.5	9
5	10	10	8
6	20	12.5	7
7	30	25	6
8	40	50	5
9	50	75	4
10	60	100	3
11	70	110	2
12	80	125	1.5
13	90	150	1
14	100	300	.75
15	110	500	.5
16	120	1000	.25

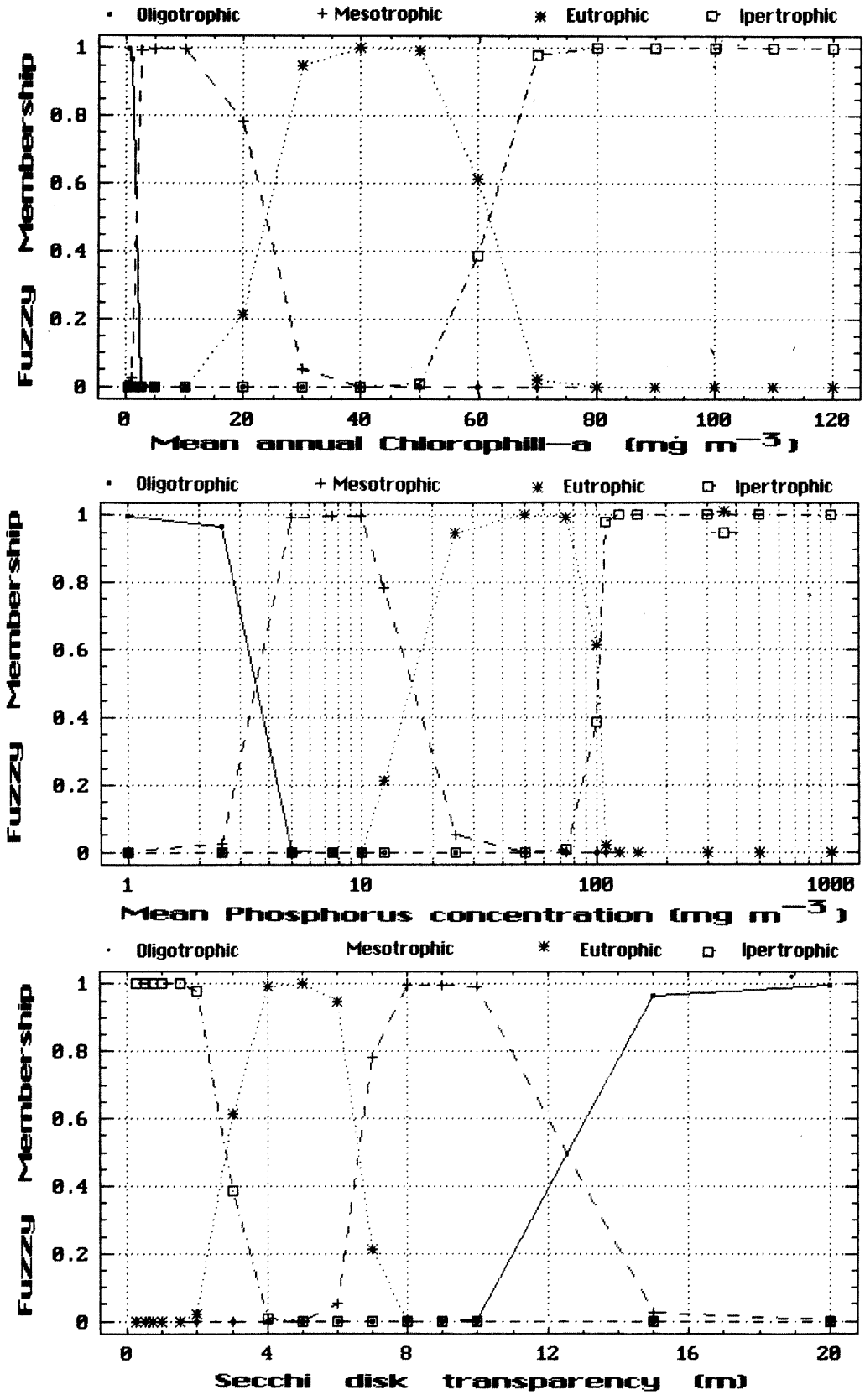

Fig. 3. Fuzzy membership functions for the eutrophication classes, generated with Algorithm 1, plotted versus set supports. a) - Chlorophyll-a; b) - Phosphorus; c) - Transparency.

Table 9. Prototypes of the eutrophication stages.

Stage	Chlorophyll-a (mg m^{-3})	Phosphorus (mg m^{-3})	Transparency (m)
Oligotrophic	0.745	1.735	17.55
Mesotrophic	8.512	8.481	8.63
Eutrophic	42.325	56.201	4.76
Ipertrophic	93.750	354.827	1.073

Fig. 4. Normalized partition entropy of the fuzzy clusters (m = 1.5) obtained from the data from Lagonegro and Feoli (1985).

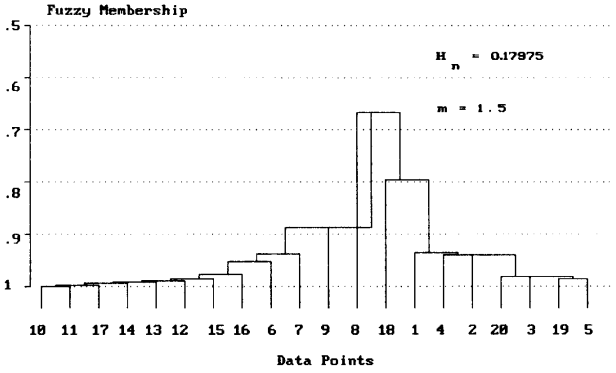

Fig. 6. Fuzzy dendrogram of the soil samples of Table 8 with fuzziness m = 1.5.

7.5 Fuzzy dendrogram of soil data

As an example of Algorithm 3 described in section 6 a fuzzy dendrogram is generated from a group of soils from a mountain area in North-East Italy (Lagonegro and Feoli 1985) characterized through chemical and meteorological parameters. The raw data are reported in Table 10. They comprise twenty samples, each encompassing eight different parameters, such as moisture, pH, nutrient content, humus, texture, sunlight exposure, temperature and thermal excursion. Applying Algorithm 1 to this data set does not result in any significant partition for any C greater than 2, as the increasing partition entropy of Fig. 4 shows. Thus dendrograms can be used to show the progressive aggregation of points. Figs. 5 and 6 compare the conventional and fuzzy dendrograms. It can be seen that in both cases samples {1, 2, 3, 4, 5, 19, 20} reveal strong similarities, since they are aggregated at a low threshold level in the deterministic case (Fig. 5) or with a nearly-one membership in the fuzzy case (Fig. 6). Another group, formed by samples {10, 11, 12, 13, 14, 15, 16, 17} also show a definite affinity, whereas in both cases points {8, 9, 18} are aggregated at the highest level. It can be seen that both approaches yield similar conclusion, with the advantage of a much smaller computational burden for the fuzzy approach.

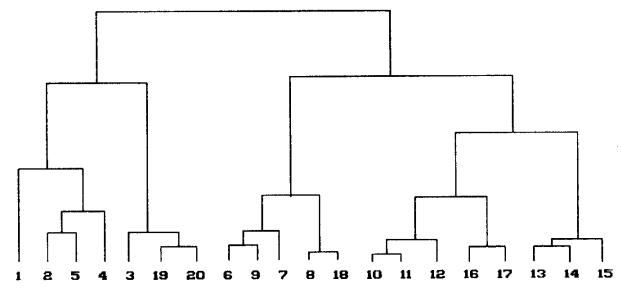

Fig. 5. Deterministic dendrograms of the soil samples of Table 8 (redrawn from Lagonegro and Feoli 1985).

8. Conclusion

This paper has presented a clustering approach based on fuzzy sets, which is deemed suitable for processing ecological data in which intermingling of underlaying structures is often impossible to discern using conventional clustering algorithms. Three algorithms were illustrated to perform ordination (Algorithm 1), classification (Algorithm 2) and fuzzy dendrogram generation (Algorithm 3) based on fuzzy concepts. The first algorithm generates a matrix of fuzzy memberships from the original data and produces representative prototypes. These can be used later to classify any subsequent data of the same type. Algorithm 2 performs this task. Finally, the generation of fuzzy dendrograms (Algorithm 3) is presented.

The main advantages of the fuzzy approach over conventional clusters are the computational simplicity and the ability to yield a graded membership of data. These features are particularly appropriate when dealing with ecological data, as shown with several ecological

Table 10. Soil samples from North-East italian mountains (Lagonegro and Feoli (1985).

Sample	M	pH	N	H	D	L	T	C
1	1.76	3.84	2.28	2.90	3.12	3.06	4.71	2.60
2	2.07	3.72	2.55	3.10	2.95	2.98	4.46	2.70
3	1.90	3.91	2.22	2.94	2.81	3.16	4.02	3.25
4	2.30	3.57	2.46	3.10	3.37	2.96	4.16	2.85
5	1.96	3.68	2.32	3.07	3.17	3.05	4.29	2.95
6	2.45	3.42	2.54	3.27	3.49	2.82	3.89	2.84
7	2.44	3.20	2.45	3.28	3.54	2.89	3.77	2.83
8	2.34	3.60	2.45	3.28	3.32	2.83	3.84	3.04
9	2.47	3.52	2.54	3.26	3.53	2.80	3.78	2.90
10	2.61	3.33	2.62	3.37	3.58	2.66	3.73	2.84
11	2.69	3.38	2.69	3.38	3.61	2.65	3.87	2.67
12	2.64	3.10	2.62	3.39	3.63	2.66	3.77	2.65
13	2.77	3.43	2.87	3.44	3.68	2.46	3.80	2.60
14	2.82	3.32	2.91	3.46	3.66	2.41	3.64	2.68
15	2.92	3.37	3.01	3.46	3.57	2.35	3.54	2.70
16	2.65	3.50	2.64	3.32	3.45	2.72	3.51	2.94
17	2.68	3.41	2.73	3.40	3.49	2.53	3.50	2.87
18	2.40	3.64	2.53	3.21	3.25	2.86	3.58	3.09
19	2.13	3.56	2.27	3.05	3.04	3.06	3.71	3.19
20	2.03	3.68	2.19	2.99	2.83	3.17	3.62	3.24

Symbols: M = Moisture; pH = soil pH; N = Nutrients; H = Humus; D = Texture; L = Light; T = Temperature; C = Thermal excursion.

examples presented in section 7 to demonstrate the algorithms.

Acknowledgements. This research was supported by CNR, Italy, under Special Grant I.P.R.A. (Productivity Enhancement of Agricultural Resources), Subproject 1, paper n. 2072.

REFERENCES

BEZDEK, J.C. 1974. Numerical Taxonomy with Fuzzy Sets. J. Math. Biol., vol. 1-1, pp. 57-71.

BEZDEK, J.C. 1981. Pattern Recognition with Fuzzy Objective Function. Plenum Press, New York, p. 378.

CORMACK, R.M. 1971. A Review of Classification. Proc. Royal Statistical Society, Vol. 3: 321-353.

DILLON, R.W. and M. GOLDSTEIN. 1984. Multivariate Analisys: Methods and applications. John Wiley & Sons, New York, p. 587.

DUNN, J.C. 1974a. A Fuzzy Relative of the ISODATA Process and its Use in Detecting Compact, Well-Separated Clusters. J. Cybern., vol. 3: 32-57.

DUNN, J.C. 1974b. Well Separated Clusters and Optimal Fuzzy Partitions in Large Data Sets, J. Cybern., Vol. 4-1: 95-104.

FEOLI, E. and V. ZUCCARELLO. 1986. Ordination based on classification: yet another solution? Abstracta Botanica 10: 203-219.

FEOLI, E. and V. ZUCCARELLO. 1988. Syntaxonomy: a source of useful fuzzy sets for environmental analysis? Coenoses 3: 141-147.

FISHER, R.A. 1936. The Use of Multiple Measurements in Taxonomic Problems, Ann. Eugenics, 7: 179-188.

GAINES, B.R. and L.J. KOHOUT. 1977. The Fuzzy Decade: A Bibliography of Fuzzy Systems and Closely Related Topics. Int. J. Man-Mach. Stud., vol. 9: 1-68.

GOWER, J.C. and G.J.S. ROSS. 1969. Minimum Spanning Trees and Single Linkage Cluster Analysis. Applied Statistics, Vol. 18: 54-64.

GUSTAFSON, D.E. and W. KESSEL. 1979. Fuzzy Clustering with a Fuzzy Covariance Matrix, Proc. IEEE-CDC, vol. 2 (K.S. Fu, ed.): 761-766, IEEE Press, Piscataway (N.J.).

INGRAM, K.T., D.C. HERZOG, K.J. BOOTE, J.W. JONES and C.S. BARFIELD. 1981. Effects of Defoliating Pests on Soybean Canopy CO_2 Exchange and Reproductive Growth, Crop Science, 21: 961-968.

JAMES, M. 1985. Classification Algorithms, Collins, London, p. 210.

KANDEL, A. 1986. Fuzzy Mathematical Techniques with Applications, Addison Wesley, Reading, Mass., p. 274.

KAUFMANN, A. and M.M. GUPTA. 1985. Introduction to Fuzzy Arithmetics: theory & applications, Van Nostrand Reinhold Computer science and Engineering Series, New York, p. 351.

KERSHAW, K.A. 1973. Quantitative and Dynamic Plant Ecology, 2nd ed., E. Arnold, London, p. 308.

LAGONEGRO, M. and E. FEOLI. 1985. Multivariate Data Analysis. (Italian), Libreria Goliardica editrice, Trieste, p. 213.

MARCHETTI, R. 1987. L'Eutrofizzazione, un processo degenerativo delle acque, Franco Angeli, Milan, p. 315 (in italian).

MARSILI-LIBELLI, S. 1985a. Mathematical Modelling of Plant Growth and Stress, Acta Horticolturae, Vol. 171: 361-370.

MARSILI-LIBELLI, S. 1985b. Parameter Identification of Plant Growth Models, Proc. 7th IFAC/IFORS Symp. on Identification and System Parameter Estimation, (edited by H.A. Barker and P.C. Young), Pergamon Press, Oxford, p. 475-462.

MARSILI-LIBELLI, S. 1986. Crop Growth and Stress Assessment through Fuzzy Clustering, Proc. 2nd European Simulation Congress (edited by P. Geril), SCS Europe Press, Ghent, p.

31-37.

NEGOITA, C.V. 1985. Expert Systems and Fuzzy Systems, The Benjamin Cummings Publ. Co., Menlo Park, Calif., p. 190.

ORLÓCI, L. 1978. Multivariate Analysis in Vegetation Research 2nd ed.). Dr. W. Junk, The Hague.

ORLÓCI, L. 1982. Numerical Methods in Ecology and Systematics, Lecture notes, London, Ontario.

PIELOU, E.C. 1984. The Interpretation of Ecological Data, John Wiley & Sons, New York, p. 263.

RUSPINI, E. 1969. A New Approach to Clustering. Inf. Control, Vol. 15: 22-32.

YAGER, R.R., R.M. OVCHINNIKOV, H.T. TONG and NGUYEN (editors). 1987. Fuzzy Sets and Applications: selected papers by L.A. Zadeh, J. Wiley & Sons, New York, p. 684.

ZHAO, S.X. 1986. Discussion on Fuzzy Clustering. 8th int. Conf. on Pattern Recognition, IEEE Press, New York, p. 612-614.

ZADEH, L. 1965. Fuzzy Sets, Information and Control, Vol. 8: 338-353.

ZIMMERMAN, H.J. 1985. Fuzzy Set Theory and its Applications, Kluwer-Nijhoff Publishing, Boston, p. 363.

Manuscript received: February 1989