Aquat. Living Resour. 12 (3) (1999) 167–175 © Ifremer/Cnrs/Inra/Ird/Cemagref/Elsevier, Paris

Species distribution in the southern Aegean sea based on bottom-trawl surveys

George Tserpes [a*], Panagiota Peristeraki [a], George Potamias [b], Nickos Tsimenides [a]

[a] *Institute of Marine Biology of Crete, P.O. Box 2214, 71003 Iraklion, Greece.*

[b] *Institute of Computer Science, P.O. Box 1385, 71110 Iraklion, Greece.*

Received July 24, 1998; accepted January 18, 1999.

Abstract — Information on the distribution of the benthopelagic fauna of the southern Aegean sea was collected through two bottom-trawl surveys carried out at fixed sampling stations in the summers of 1996 and 1997. Discriminant analysis identified that the most important species in discriminating among station-groups in terms of numbers were: *Pagellus erythrinus*, *Mullus barbatus*, *Mullus surmuletus*, *Diplodus annularis*, *Spicara flexuosa*, *Sardina pilchardus* and *Parapenaeus longirostris*. In terms of weight, the most important discriminating species were: *Mullus barbatus*, *Pagellus erythrinus*, *Spicara maena*, *Scyliorhinus canicula*, *Loligo vulgaris* and *Parapenaeus longirostris*. Supervised machine learning approaches and, in particular, the decision tree construction method were utilized in order to induce rules which determine the station-grouping. Several species, notably, *Argentina sphyraena*, *Aristaeomorpha foliacea*, *Lepidorhombus bosci*, *Lepidotrigla cavillone*, *Mullus barbatus*, *Serranus cabrilla* and *Sepia officinalis* appeared in most of the rules. © Ifremer/Cnrs/Inra/Ird/Cemagref/Elsevier, Paris

Bottom-trawl surveys / distribution / multivariate analysis / machine learning / Mediterranean sea

Résumé — Deux campagnes de chalutage de fond ont été menées, en été 1996 et 1997, sur une série de stations déterminées. Les informations ont été ainsi fournies sur la répartition de la faune benthique et bentho-pélagique du Sud de la mer Egée. Une analyse statistique discriminante a permis d'identifier les espèces les plus importantes en fonction du regroupement des stations. Les espèces discriminantes les plus importantes sont : (a) en nombre, *Pagellus erythrinus*, *Mullus barbatus*, *Mullus surmuletus*, *Diplodus annularis*, *Spicara flexuosa*, *Sardina pilchardus* et la crevette *Parapenaeus longirostris* ; (b) en poids, *Mullus barbatus*, *Pagellus erythrinus*, *Spicara maena*, *Scyliorhinus canicula*, l'encornet *Loligo vulgaris* et la crevette *Parapenaeus longirostris*. Des approches au moyen de « méthodes d'apprentissage automatique supervisé » et, en particulier d'arbres de décision, ont été utilisées afin d'établir les règles de décision déterminant le regroupement des stations. Plusieurs espèces apparaissent dans la plupart des règles de décision de ces regroupements, en particulier : *Argentina sphyraena*, *Aristaeomorpha foliacea*, *Lepidorhombus bosci*, *Lepidotrigla cavillone*, *Mullus barbatus*, *Serranus cabrilla* et *Sepia officinalis*. © Ifremer/Cnrs/Inra/Ird/Cemagref/Elsevier, Paris

Chalutage de fond / répartition des espèces / analyse multivariée / arbre de décision / mer Méditerranée

1. INTRODUCTION

Information on the macrofauna of the southern Aegean sea is limited to a few taxonomic reports [1, 6, 13, 17] and studies on the community structure of the benthopelagic species inhabiting the area are scanty and confined to small sub-areas. The presence of numerous islands gives some unique topographical and bathymetrical features to the area which contribute to its zoogeographical distinction from the rest of the Aegean [12].

The present work uses data from a bottom-trawl survey on the marine species of the southern Aegean sea, carried out within the framework of an international Mediterranean project and is mainly intended to describe broadly the benthopelagic fauna of the region and provide baseline information for further studies in the area. Data are analysed by means of: (i) traditional

* Corresponding author, e-mail: gtserpes@crete.cc.uch.gr

multivariate techniques such as cluster and discriminant analysis and (ii) up-to-date supervised machine learning approaches and in particular the inductive decision tree construction method. The decision tree method has not been previously used for the analysis of such data and it is applied in order to identify species composition patterns which govern the classification of the sampling stations into groups. The method has the advantage of being free from any distributional assumption implicit in multivariate analysis. The suitability of the decision tree method is discussed and its results are compared with those of the discriminant analysis.

2. MATERIALS AND METHODS

2.1. Sampling

Two experimental bottom-trawl surveys were carried out in the southern Aegean sea, in the area extending southern of the 38th parallel down to the Cretan coast, during the summer of 1996 and 1997. The 1996 survey was carried out from the beginning of June until the middle of July while the 1997 survey started at the beginning of July and finished in the middle of August. The surveys were carried out within the framework of the Mediterranean International Trawl Survey (MEDITS) project and aimed at collecting data for a reference list of fish, cephalopod and crustacean species at fixed sampling stations [2]. The stations were selected on a depth-stratified random design in a way that permitted us to sample the following depth intervals: 0–50, 50–100, 100–200, 200–500 and 500–800 m. In the southern Aegean sea, 111 bottom-trawls were taken, 51 in 1996 and 60 in 1997 (*figure 1*). Due to bad weather conditions, nine of the trawls were not taken in both years. The duration of each trawl varied from 30–60 min. The trawling speed was 2.5 knots. Since all trawls were carried out using the fishing vessel 'Ioannis Rossos' and the same fishing gear, it was assumed that gear selectivity was constant.

2.2. Data analysis

For each haul, the door spread of the trawl was multiplied by the speed of the boat and fishing time in order to obtain the total sampled area. This procedure enabled us to express the species number and weight, in number and kg per square kilometer (km^2) respectively, and made possible the comparisons between stations. As very few hauls (two per year) were made in the 0–50-m depth interval, it was added to the 50–100 interval creating the 0–100-m one. Therefore, we had all sampling stations belonging in the following four depth intervals: 0–100, 100–200, 200–500 and 500–800 m.

Differences in distribution patterns of fish between stations both in terms of numbers and weight were determined separately for each cruise using cluster analysis. The mean Euclidean distance (MED) was used to represent the dissimilarity between stations and clustering was done using the average method.

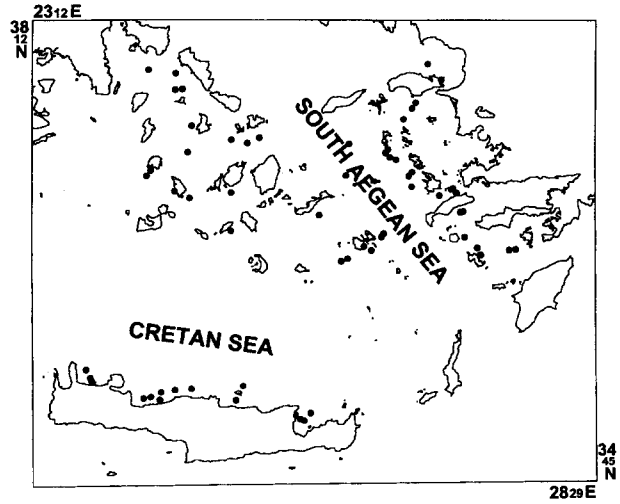


Figure 1. Sketch of the area indicating the sampling stations.

In order to determine if the station-groups defined by the cluster analysis differed significantly in terms of species occurrence and to define the species which were the most important for differentiating among groups, discriminant analysis was applied using as variables: (i) the number and (ii) the weight of the species. The analysis was applied for each survey separately. The distinctiveness of station-groups was measured using: (i) the Wilks' λ-criterion and its correspondent F-statistic to test the significance of the overall difference between group centroids [16, 20] and (ii) the squared canonical correlation for each discriminant function which was interpreted as the part of the total variance in the corresponding discriminant function accounted for by the groups [9]. Another indirect measure of the adequacy of the classification feature was the percentage of stations correctly assigned to groups. The standardized discriminant function coefficients assigned to the variables by the discriminant analysis were interpreted as giving the relative importance of each variable in separating station-groups [8, 20].

Supervised machine learning approaches [11], in particular the decision tree construction method [14], were applied to identify species composition patterns which govern the classification of the stations into the pre-defined groups (classes). The method relies on information theoretic measures, namely the 'Entropy' metric. Each feature (species number or weight in our case) is ranked according to its 'informative power' to distinguish between the classes assigned to the samples of the current data subset. Entropy measures the 'disorder' present in the data set with respect to distinguishing between classes. For example, in a

Table I. Occurrence by depth zone of the species for which data have been collected in the southern Aegean sea.

Species name	Code name	1996 Depth zone (m)				1997 Depth zone (m)			
		0–100	100–200	200–500	500–800	0–100	100–200	200–500	500–800
Argentina sphyraena	ARGESPY		X	X	X		X	X	X
Aristeomorpha foliacea	ARISFOL				X				X
Aspitrigla cuculus	ASPICUC		X	X		X	X	X	
Boops boops	BOOPBOO	X	X	X		X	X	X	
Citharus linguatula	CITHMAC	X	X			X	X	X	
Conger conger	CONGCON	X		X		X	X	X	
Diplodus annularis	DIPLANN	X				X			
Eledone cirrosa	ELEDCIR			X		X	X	X	X
Eledone moschata	ELEDMOS	X	X			X	X	X	
Engraulis encrasicholus	ENGRENC					X			
Helicolenus dactylopterus	HELIDAC	X	X	X	X		X	X	X
Illex coindetii	ILLECOI	X	X	X		X	X	X	X
Lepidopus caudatus	LEPICAU	X	X	X	X		X	X	
Lepidorhombus boscii	LEPMBOS	X		X	X			X	X
Lepidotrigla cavillone	LEPTCAV	X	X	X		X	X	X	
Loligo vulgaris	LOLIVUL	X	X			X	X		
Lophius budegassa	LOPHBUD	X	X	X	X	X	X	X	X
Lophius piscatorius	LOPHPIS		X	X			X	X	
Merluccius merluccius	MERLMER	X	X	X	X	X	X	X	X
Micromesistius poutassou	MICMPOU			X	X			X	X
Mullus barbatus	MULLBAR	X	X	X		X	X	X	
Mullus surmuletus	MULLSUR	X	X	X		X	X	X	
Nephrops norvegicus	NEPRNOR		X	X	X		X	X	X
Octopus vulgaris	OCTOVUL	X	X	X		X	X	X	
Pagellus acarne	PAGEACA	X	X	X		X	X	X	
Pagellus bogaraveo	PAGEBOG			X	X	X		X	X
Pagellus erythrinus	PAGEERY	X	X			X	X	X	
Parapenaeus longirostris	PAPELON	X	X	X	X	X	X	X	X
Phycis blennoides	PHYIBLE		X	X	X		X	X	X
Raja asterias	RAJAAST		X	X					
Raja clavata	RAJACLA	X	X	X	X	X	X	X	X
Sardina pilchardus	SARDPIL	X	X			X	X		
Scorpaena notata	SCORNOT	X				X	X	X	
Scyliorhinus canicula	SCYOCAN	X	X	X	X	X	X	X	X
Sepia officinalis	SEPIOFF	X	X			X	X		
Sepia orbignyana	SEPIORB	X	X	X	X	X	X	X	
Serranus cabrilla	SERACAB	X	X			X	X		
Spicara flexuosa	SPICFLE	X	X			X	X		
Spicara maena	SPICMAE	X		X		X	X		
Spicara smaris	SPICSMA	X	X	X		X	X	X	
Squalus acanthias	SQUAACA		X	X	X		X	X	X
Trachurus mediterraneus	TRACMED	X	X	X		X	X	X	X
Trachurus picturatus	TRACPIC	X		X	X				
Trachurus trachurus	TRACTRA	X	X	X	X	X	X	X	
Trigla lucerna	TRIGLUC	X				X		X	
Trisopterus minutus capelanus	TRISCAP		X	X		X	X	X	
Zeus faber	ZEUSFAB	X	X	X		X	X	X	

two-classes data set E with p_i, n_i samples belonging to classes p, n, respectively, the entropy is computed by the following formula:

$$Entropy = -\sum \frac{p_i}{p_i + n_1} \log \frac{p_i}{p_i + n_i} - \sum \frac{n_i}{p_i + n_1} \log \frac{n_i}{p_i + n_i}$$

The 'Conditional Entropy', i.e. the present disorder in the data set when a particular feature is selected, is computed by the following formula:

$$Cond.\ Entropy = \sum_{value} \frac{|E[Feature\ (Value)]|}{|E|} \times Entropy\ (E[Feature\ (Value)])$$

Table II. Results of the discriminant analysis (only indices corresponding to the first two more important functions have been included).

Year	Wilks' λ (χ^2-test)	Discriminant function	Percentage of variance	Squared canonical correlation	Correct assignment (%)
1996 (number)	$3 \cdot 10^{-5}$ ($P << 0.001$)	1	61	0.99	100
		2	33	0.98	
1996 (weight)	$8 \cdot 10^{-5}$ ($P << 0.001$)	1	53	0.98	100
		2	39	0.97	
1997 (number)	10^{-4} ($P << 0.001$)	1	74	0.98	100
		2	18	0.94	
1997 (weight)	$2 \cdot 10^{-4}$ ($P << 0.001$)	1	73	0.98	100
		2	14	0.91	

where, *E*[*Feature*(*Value*)] is the subset of the data in which all examples share value *Value* for feature *Feature*, [*E*] the cardinal of data set *E* (= number of samples).

Finally, it is natural to assume that a 'good feature' is the one that 'minimizes' the disorder of the data set when it is selected. In other words, the most informative feature is the one with the highest computed score for the formula:

InfoGain(Feature) = Entropy – Conditional Entropy

which computes the 'Information Gain' of a feature.

The way that the tree is constructed (induced) is based on a 'divide and conquer' strategy. Within this approach, the given set of data is partitioned into subsets and for each of the subsets, the 'most informative' feature is identified. Following a 'hill-climbing' approach and iteratively dividing the data set into subsets by selecting the 'most informative' feature, the whole tree is constructed. The inductive decision tree construction process could be considered as a divisive clustering approach where the data set is partitioned and each partition incorporates similar samples. The whole approach falls into the category of symbolic data analysis [5].

Prior to the analysis, the log(x + 1) transformation was used. The logarithmic transformation commonly used in this kind of data minimizes the effect of anomalous catches and allows greater contribution from the rarer species [3, 4]. The addition of one unit was necessary to avoid problems derived by the presence of zero values. All statistical inferences were based on the 0.05 significance level.

The cluster and discriminant analysis were performed by means of the statistical package Systat [22]. For the decision tree analysis, the package C4.5 [15] was used.

3. RESULTS

Abundance data (number and weight) were collected for the 47 species found out of the 58 in the project's reference list (*table I*). Cluster analysis showed that the major station-clusters were generally delimited by the sampled depth intervals (*figure 2*).

Results of the discriminant analysis are shown in *table II*. As shown by the Wilks' λ-criterion, the differences among station-groups were always highly significant ($P << 0.001$; all comparisons). In each case, analysis yielded three discriminant functions but, to save space, only indices corresponding to the first two most important functions have been included. The influence of station-groups was highly substantial as indicated by the squared canonical correlation values. They indicated that station-groups always accounted for more than 90 % of the variation among sampling stations with respect to the first two discriminant functions. Moreover, the 100 % correct assignments of stations to the groups indicated satisfactory discrimination between groups and a plot of the discriminant scores for the samples on the discriminant space revealed a clear separation of station-groups (*figures 3, 4*). The first discriminant function, especially in 1997, was much more important than the second in separating station-groups as indicated by the percentage of the explained variance values.

The standardized discriminant function coefficients allowed us to identify the more important species in discriminating among station-groups. The main discriminating species as revealed by the analysis differed according to the variable used (number or weight). This is expected as in general the variables were not analogous, i.e. high number of a certain species in one station did not necessarily imply high biomass values and vice-versa. *Loligo vulgaris* was the only species that appeared with relatively high coefficients, both in terms of number and weight, in all years. In terms of numbers, *Eledone moschata* had high coefficients in both examined years while, in terms of weight, the same was valid for *Mullus barbatus*, *Scyliorhinus canicula* and *Spicara flexuosa*. The ten most important species in discriminating among station-groups with respect to the first discriminant function are shown in *table III*.

The decision tree construction method revealed a series of rules relating the depth location of the stations with species composition. *Table IV* shows the main rules identified for each data set and the percentage of success they had in classifying the stations into the corresponding depth group. To save space, only rules having a discrimination success over 85 % are

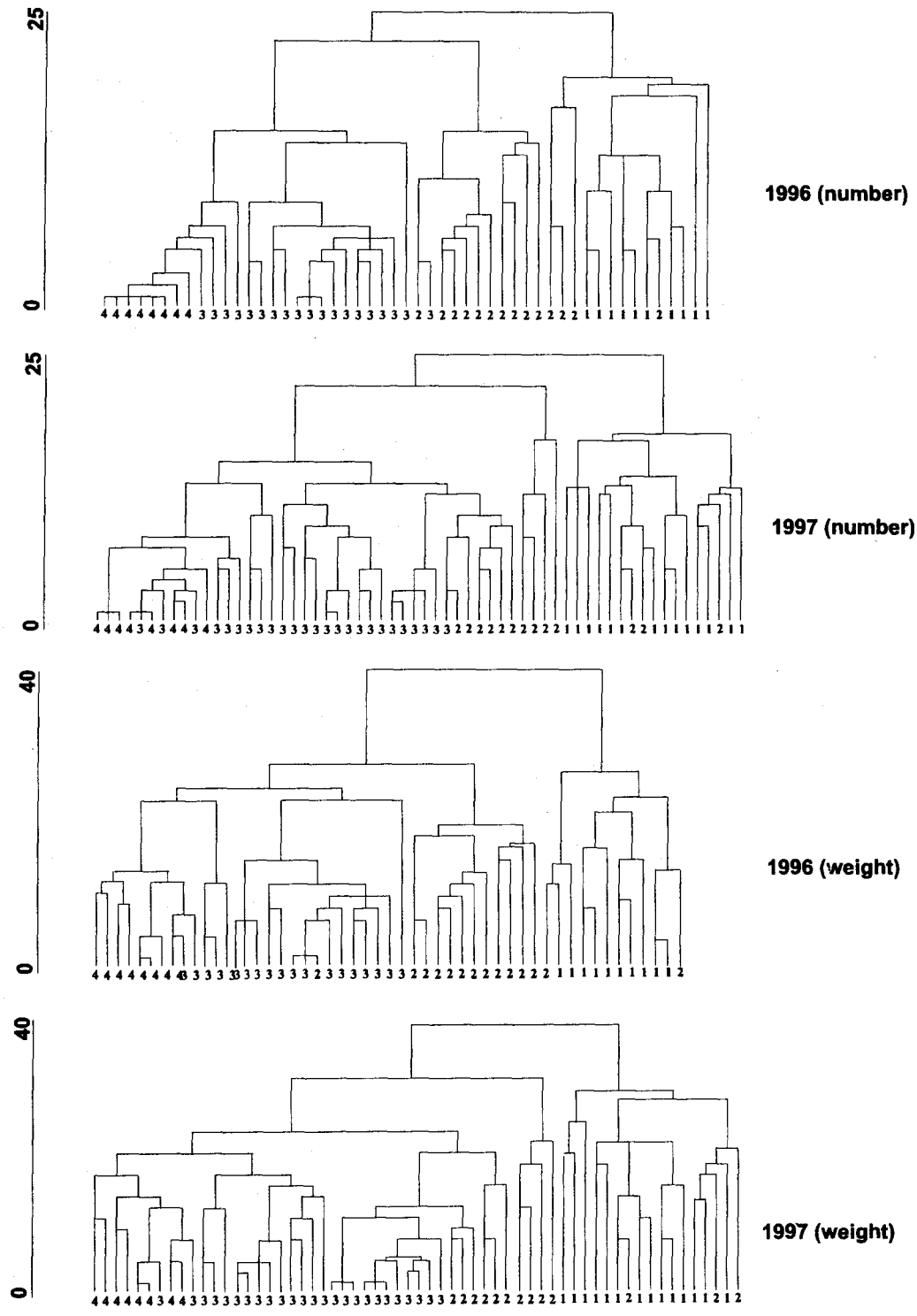
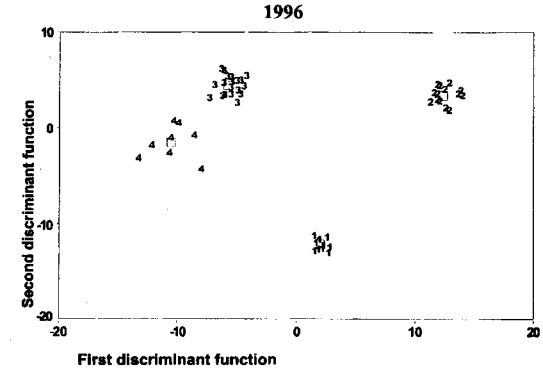


Figure 2. Station clustering for both the 1996 and 1997 cruises based on the analysis of species number and weight. Vertical lines indicate the percentage dissimilarity between stations. Numbers indicate the depth-group where the corresponding station belongs: 1 = 0–100, 2 = 100–200, 3 = 200–500 and 4 = 500–800 m.



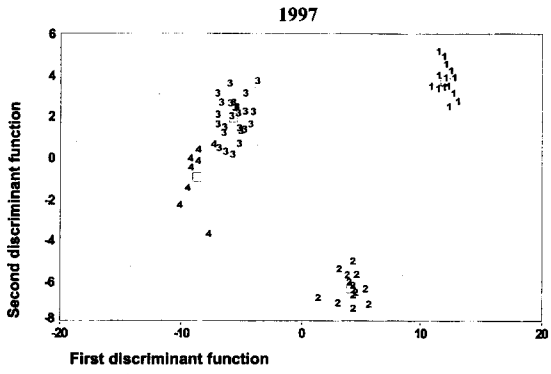
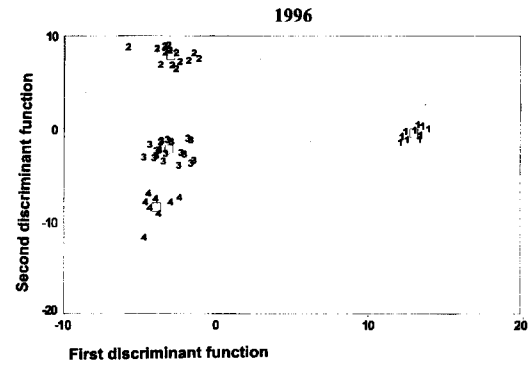


Figure 3. Discriminant space for both the 1996 and 1997 cruises based on the analysis of species abundance in terms of numbers. Each number represents a sampling station with a value equal to the depth-group where the station belongs: 1 = 0–100, 2 = 100–200, 3 = 200–500 and 4 = 500–800 m. Squares indicate the station-group centroids.



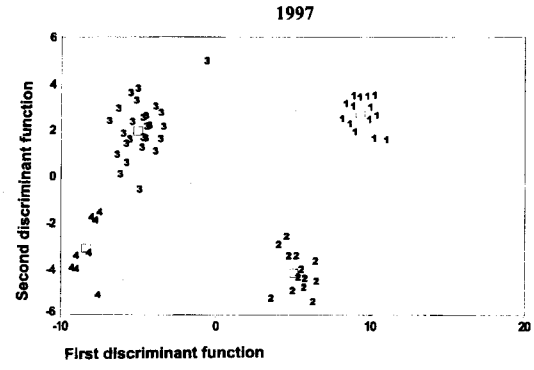


Figure 4. Discriminant space for both the 1996 and 1997 cruises based on the analysis of species abundance in terms of weight. Each number represents a sampling station with a value equal to the depth-group where the station belongs: 1 = 0–100, 2 = 100–200, 3 = 200–500 and 4 = 500–800 m. Squares indicate the station-group centroids.

included in the table and, as it can be seen, most of these rules deal with the 200–500-m depth zone. The rules made possible the identification of some species relationships which characterize the examined depth zones. The main rules identified, differed from one year to another, but the fact that certain species, such as *Argentina sphyraena*, *Aristeomorpha foliacea*, *Lepidorhombus bosci*, *Lepidotrigla cavillone*, *Mullus barbatus*, *Serranus cabrilla* and *Sepia officinalis*, appeared in rules of different data sets suggests that these species play an important role in discriminating among station-groups.

The discrepancies concerning the main discriminating species between the 1996 and 1997 surveys can be

Table III. Standardized discriminant function coefficients (in parentheses), with respect to the first discriminant function, for the ten most important species in discriminating among station-groups. Code names correspond to the species names as mentioned in *table I*.

Year	Species
1996 (number)	DIPLANN (10.99), SCORNOT (–7.47), SPICFLE (5.49), ELEDMOS (–5.20), PAGEERY (–4.52), TRIGLUC (–4.48), LEPTCAV (3.80), SEPIORB (–3.13), CITHMAC (2.99), LOLIVUL (2.83)
1996 (weight)	SCYOCAN (7.39), PAGEERY (6.35), SPICMAE (5.97), BOOPBOO (–5.44), OCTOVUL (–5.33), LOLIVUL (5.28), LOPHBUD (–4.23), SPICFLE (3.94), MULLBAR (3.18) ELEDCIR (–3.29)
1997 (number)	MULLSUR (–3.74), SARDPIL (2.95), PAPELON (2.86), MULLBAR (2.82), ENGRENC (–2.53), LOLIVUL (2.47), ZEUSFAB (2.22), SERACAB (1.64), TRACTRA (–1.53), ELEDMOS (1.23)
1997 (weight)	PAPELON (2.99), ENGRENC (–2.51), ARGESPY (–2.46), SQUAACA (2.02), MULLBAR (1.91), SCYOCAN (1.72), LOLIVUL (1.69), SPICFLE (1.55), SARDPIL (1.51), ASPICUC (1.50)

Table IV. The main rules identified by the decision tree method. Numbers in parentheses correspond to the percentage classification success of each rule. The last column indicates the overall classification success with respect to all rules identified for the specific data set. Code names correspond to the species names as mentioned in *table I.*

Year	Rule		Overall success
	Condition (expressed in numbers or kg·km^{-2})	Station depth location (m)	
1996 (number)	ARISFOL = 0, LEPMBOS > 0, LOLIVUL = 0	200–500 (90.6 %)	88.2 %
1996 (weight)	ARISFOL = 0, LEPMBOS > 0, LEPTCAV ≤ 15.03	200–500 (90.6 %)	86.3 %
	ARGESPY > 0, LEPTCAV > 15.03	100–200 (87.1 %)	
	ASPICUC > 0, LEPTCAV > 15.03	100–200 (85.7 %)	
1997 (number)	LEPICAU > 0, MULLBAR ≤ 6.62	200–500 (92.2 %)	91.7 %
	ARGESPY > 3.41, SERACAB = 0	200–500 (98.9 %)	
	ARGESPY = 0, SEPIOFF ≤ 5.81, SERACAB > 0	0–100 (88.2 %)	
1997 (weight)	RAJACLA > 40.85, SARDPIL = 0, SERACAB = 0, SPICFLE = 0	200–500 (93.6 %)	100 %
	MULLBAR = 0, SEPIORB > 0, SERACAB = 0, SPICFLE = 0	200–500 (91.7 %)	
	ARGESPY = 0, ELEDMOS > 44.70	0–100 (87.1 %)	
	ARGESPY = 0, SEPIOFF = 0, SERACAB > 0	0–100 (85.7 %)	

attributed to the one-month difference between the two surveys. As the most common species (e.g. Sparidae, Mullidae, etc.) are recruited into the fishery during the summer months, it is normal to expect that the sampling-time difference affected the measured variables (number and weight) and consequently the results of both the dicriminant and decision tree analysis. The abundance cut-off points that have been identified by the rules are always very close to the lower tail of the corresponding frequency distributions of abundance indicating the strong influence of the stratum boundaries in the distribution of the species.

4. DISCUSSION

Our results provided a quantitative basis for the division of the benthopelagic fauna of the southern Aegean sea into components according to depth. The main features which discriminate the 0–100-m zone are the existence of *Diplodus annularis* and the relatively higher abundance of *Eledone moschata*, *Loligo vulgaris* and *Octopus vulgaris*. There is a group of species such as *Pagellus erythrinus*, *Serranus cabrilla*, *Loligo vulgaris*, *Eledone moschata* and *Spicara flexuosa* exclusively found in both the 0–100 and 100–200-m zones and few more species such as *Boops boops*, *Lepidotrigla cavillone*, *Mullus surmuletus* and *Citharus linguatula* which are found in relatively higher abundance in those depths. All the aforementioned species could be considered as the dominant species of the southern Aegean continental shelf which usually extends to depths close to 200 m.

The 200–500 and 500–800-m zones are characterized by the relatively higher abundance of a group of deep-water species such as: *Phycis blennoides*, *Squalus acanthias*, *Nephrops norvegicus* and *Parapenaeus longirostris*. However, the 200–500-m zone has also similarities with the former 100–200 as species inhabiting the end of the shelf, such as *Argentina sphyraena*, *Aspitrigla cuculus*, *Lophius budegassa*, *Scyliorhinus canicula* and *Sepia orbignyana*, are mostly found in these two zones. The only species found exclusively in the 500–800-m zone was the red shrimp *Aristeomorpha foliacea.*

All the aforementioned species were revealed by the analysis to be important in discriminating among station-groups. Furthermore, a few additional species made high contributions to the discriminant functions and/or were identified as important by the decision tree method. This included some pelagic species such as *Engraulis encrasicholus* and *Sardina pilchardus* but also demersal species such as *Mullus barbatus*, *Lepidopus caudatus*, *Lepidorhombus boscii* and *Zeus faber.* The above demersal species had a wide depth distribution, which in the case of *Lepidopus caudatus* and *Lepidorhombus boscii* covered all the examined depth intervals. However, these species were identified as main discriminating variables because they inhabited the different depths in different densities. For instance, *Mullus barbatus* which inhabited all depths from 0–500 m was three to five times more abundant in the 0–100-m zone than in the 200–500 one.

In the past, different multivariate techniques have been used for the analysis of trawl survey data [3, 10, 21, 24] but, to the best of our knowledge, no attempt has been made to analyse such data by means of machine learning approaches. A somewhat similar approach which included a binary decision tree has been utilized for the definition of the typology of a trawler fleet in the south of France [19]. In the present case however, we followed a non-binary decision tree construction approach with multi-valued splits per discriminant attribute which gave the potential for discovering different discriminant patterns that characterize each station-group and revealed multiple rules.

Recently, machine learning approaches have gained an increasing interest, both academic and commercial, in data analysis operations. We refer to the innovative data mining technology [7], which seems to be the

state-of-the-art in large data analysis and knowledge discovery from databases. In the present case, the decision tree method proved to be quite effective in the direct identification of species groups characterizing the different depth zones.

Even though the decision tree method cannot detect the relative importance of each discriminating feature, as does the discriminant analysis, it has the advantage of being free from the assumptions implicit in linear discrimination. These assumptions, which include data normality, homogeneous dispersions and identifiable parameter values, are difficult to test and their violation can affect the classification rates [23]. However, if (i) the statistical differences between groups are highly significant, (ii) the discriminant function coefficients are ecologically interpretable, and (iii) there are obvious species separations on each discriminant function, then it is reasonable to conclude that the identified species distribution differences represent a real situation [8]. As all the above statements are valid in the present study, our results of the discriminant analysis can be considered as indicative of the population distribution in the area.

The discrepancies in the main discriminating species between the two methods can be attributed to the fact that the decision tree method does not identify discriminating species but discriminating rules which include species groups. Naturally, it is normal to conclude that the species included in the rules are the main discriminating species but this approach is completely different from those followed in the traditional discriminant analysis.

In the present study, both methods gave results which were in general ecologically interpretable. The agreement of the methods in terms of main discriminating species was found to be higher in cases where the overall success of the decision tree method was over 90 % as it happened with the 1997 data. The decision tree method seems to be advantageous in identifying the discriminating importance of uncommon species, such as the red shrimp *Aristeomorpha foliacea*, which although exclusively found in the 500–800-m zone, was not identified as a main species by the traditional discriminant analysis, probably due to its occurrence in low abundance. Further application of both methods in other data sets would allow to check the validity of the above statements.

In any case, the decision tree method seems to be a promising tool for the identification of spatial and temporal differences in the community structure. As it has the advantage of being free from any distributional assumption of the data, it could be applied in a wide range of cases. For instance, it could be used to examine the differences between areas or the effects of anthropogenic activities. Utilization of the produced rules in specific commercial fisheries would allow identification of the origin of the catch of the fishing boats and therefore support the fishery control of the area under study.

The present study is the first in the eastern Mediterranean which attempts to describe broadly on a quantitative basis the benthopelagic fauna in a large area. Similar studies made in the past were confined to smaller areas such as, the Cretan shelf [21] and a coralligenous shelf of northern Israel [18]. Comparable findings regarding the depth distribution of several fish species examined in the present study, such as *Argentina sphyraena*, *Mullus barbatus*, *Mullus surmuletus* and *Serranus cabrilla*, have been also reported for the Cretan shelf [21] and for the northwestern Mediterranean [10]. However, direct comparisons with findings from different studies are rather difficult as the sampling schemes differ. In the present case, the attempt to cover such a large area sacrificed the degree of resolution that is desirable for a detailed description of the community characteristics but this was inevitable given the scope of the survey programme. Further studies may show that the broad trends currently identified in the southern Aegean sea represent functional relationships between the species of the marine fauna.

Acknowledgments

We wish to thank three anonymous referees for valuable comments on an earlier version of the manuscript. This work was partially funded by the European Union (DGXIV). It does not necessarily reflect the views of the Commission and in no way anticipates any future opinion of the commission.

REFERENCES

[1] Bellog G., Inventory of the fisheries resources of the Greek waters, Appendix B: Catalogue of the resources of Greek waters, Pisces 1948, 64 p.

[2] Bertrand J., Gil De Sola L., Papaconstantinou C., Relini G., Souplet A., An international bottom-trawl survey in the Mediterranean: the MEDITS programme, ICES Annual Conference, CM 1997/Y:03, 1997, 16 p.

[3] Bianchi G., Demersal assemblages of the continental shelf and upper slope of Angola, Mar. Ecol. Prog. Ser. 81 (1992) 101–120.

[4] Clarke K.R., Non-parametric multivariate analyses of

changes in community structure, Austr. J. Ecol. 18 (1993) 117–143.

[5] Diday E., Lechevallier Y., Schader M., Bertrand P., Burtschy B., New Approaches in Classification and Data Analysis, Springer-Verlag, Berlin, 1994, 693 p.

[6] Ege V., Paralepidae I (*Paralepsis* and *Lestidium*). Taxonomy, ontogeny, phylogeny, and distribution, Dana Rep. 40, 1953, 185 p.

[7] Fayyad U.M., Piatetsky-Shapiro G., Smyth P., Uthurusamy R., Advances in Knowledge Discovery and Data Mining, AAI Press/The MIT Press, Cambridge, MA, 1996, 611 p.

[8] Green R.H., A multivariate statistical approach to the Hutchinsonian niche: Bivalve molluscs of central Canada, Ecology 52 (1971) 543–556.

[9] Lebart L., Morineau A., Warwick K., Multivariate Descriptive Statistical Analysis, John Wiley, New York, 1984, 231 p.

[10] Massuti E., Renones O., Carbonell A., Oliver P., Demersal fish communities exploited on the continental shelf and slope off Majorca, Vie Milieu 46 (1996) 45–55.

[11] Mitchell T.M., Machine Learning, McGraw-Hill, New York, 1997, 414 p.

[12] Papaconstantinou C., The spreading of Lessepsian fish migrants into the Aegean Sea (Greece), Sci. Mar. Barcelona 54 (1990) 313–316.

[13] Papaconstantinou C., Tsimenidis N., Some uncommon fishes from the Aegean Sea, Cybium 7 (1979) 3–14.

[14] Quinlan R.J., Induction of decision trees, Machine Learning 1 (1986) 81–106.

[15] Quinlan R.J., C4.5: Programs for Machine Learning, Morgan Kaufmann, San Mateo, CA, 1993, 302 p.

[16] Rao C.R., Linear Statistical Inference and its Applications, 2nd ed., John Wiley, New York, 1973, 656 p.

[17] Sinis A.I., Koukouras A.S., New information on the Aegean Sea ichthyofauna, Cybium 19 (1995) 421–424.

[18] Spanier E., Pisanty S., Tom M., Almog-Shtayer G., The fish assemblage on a coralligenous shallow shelf off the Mediterranean coast of northern Israel, J. Fish Biol. 35 (1989) 641–649.

[19] Taquet M., Gaertner J.C., Bertrand J., Typologie de la flottille chalutière du port de Sète par une méthode de segmentation, Aquat. Living. Resour. 10 (1997) 137–148.

[20] Tatsuoka M., Multivariate analysis: Techniques for Educational and Psychological Research, John Wiley, New York, 1971, 215 p.

[21] Tsimenides N., Tserpes G., Machias A., Kallianiotis A., Distribution of fishes on the Cretan shelf, J. Fish Biol. 39 (1991) 661–672.

[22] Wilkinson L., Systat: The System of Statistics, Evanston, IL, 1988, 822 p.

[23] Williams B.K., Some observations on the use of discriminant analysis in ecology, Ecology 64 (1983) 1283–1291.

[24] Yoshiyama R.M., Holt J., Holt S., Godbout R., Wohlschlag E., Abundance and distribution patterns of demersal fishes on the South Texas outer continental shelf: a statistical description, Cont. Int. Mar. Sci. 25 (1982) 61–84.