Species identification of small pelagic fish schools by means of hydroacoustics in the Eastern Mediterranean Sea

TSAGARAKIS K.	1. Institute of Marine Biological Resources and Inland Waters, Hellenic Centre for Marine Research 2. Department of Biology, University of Crete, Heraklion, Greece
GIANNOULAKI M.	Institute of Marine Biological Resources and Inland Waters, Hellenic Centre for Marine Research.
PYROUNAKI M.	Hellenic Centre for Marine Research, Institute of Marine Biological Resources and Inland Waters
MACHIAS A.	Institute of Marine Biological Resources and Inland Waters, Hellenic Centre for Marine Research.

http://dx.doi.org/10.12681/mms.799

To cite this article:

Research Article

Species identification of small pelagic fish schools by means of hydroacoustics in the Eastern Mediterranean Sea

K. TSAGARAKIS[1,2], M. GIANNOULAKI[1], M. PYROUNAKI[1] and A. MACHIAS[1]

[1] Hellenic Centre for Marine Research, Institute of Marine Biological Resources and Inland Waters, Greece
[2] Department of Biology, University of Crete, Heraklion, Greece

Corresponding author: kontsag@hcmr.gr

Handling Editor: Peter Psomadakis

Abstract

Reliable biomass estimates by means of hydroacoustics largely depend on the correct identification of acoustic targets. Data collected during five summer acoustic surveys (2004-2008) in the North Aegean Sea (Greece) were analyzed to explore effective discrimination of small pelagic fish schools according to the species they belong. Discriminant Function Analyses (DFA) using bathymetric, energetic and morphometric school descriptors as explanatory variables were applied per research cruise as well as to pooled data from all surveys. Results revealed that the schools can be successfully classified into the five species considered (anchovy *Engraulis encrasicolus,* sardine *Sardina pilchardus*, round sardinella *Sardinella aurita*, blue whiting *Micromessistius poutassou,* and Atlantic mackerel *Scomber scombrus*). The percentage of correct classifications in terms of number of schools was higher in the analyses of the annual cruises (75.6%-95.36%) than in the analysis of pooled data (≈72%). This is because of (i) the lower number of species, as well as (ii) the reduced intraspecific variability, occurring in each separate cruise. Significant differences were detected among school descriptors for the different species, revealing discrete aspects of schooling behaviour for each species. The benefit of the specific approach is that the classification functions of the DFAs can be used to classify a larger set of schools, which has not been possible to assign to specific species. Overall the approach constitutes an objective, more automated and less time consuming procedure for the analysis of acoustic data and can contribute to the improvement of biomass estimates in the area.

Keywords: Schooling behaviour, hydroacoustics, school descriptors, discriminant function analysis, MEDIAS survey.

Introduction

The acoustic methodology, supported by a continuous technological development during the last decades, is an important tool for the study of small pelagic fish distribution and behaviour, constituting in parallel the most common method for their biomass estimates (Fréon & Misund, 1999; Simmonds & MacLennan, 2005). The accuracy of these estimates relies greatly on the correct identification of acoustic targets (backscattered echo from individual fish, fish shoals, schools or abiotic and non-fish living material) (Horne, 2000; Lawson *et al.*, 2001; Simmonds & MacLennan, 2005).

Accurate biomass estimates constitute important input information, among others, for stock assessment (e.g. SG-MED, 2009), individual based and food-web (Plagányi, 2007) models. The development of the Ecosystem Approach to Fisheries (Garcia *et al.*, 2003) during the last years has broadened the application field of biomass estimates for marine populations. Management measures, such as marine spatial planning and the application of fisheries restrictions is also largely based on such information, which further highlights the need for reliable biomass estimates. During the last decade survey based information is often used beyond stock assessment purposes, for assessing the spatial distribution of the target species and the study of essential fish habitats (Turner *et al.*, 2003; Valavanis *et al.*, 2008; Giannoulaki *et al.*, 2013).

Estimating abundance or biomass in acoustic surveys largely depends on allocating fish species to the echo traces and the measured acoustic backscatter. Since the validation of all the recorded targets is not possible, the acoustic methodology largely relies on ground-truthing of sampled targets through experimental fishing in order to extend the classification to all targets (Simmonds & MacLennan, 2005). Horne (2000) characterized automated classification of targets as the "Holy Grail" of fisheries acoustics, however, despite the steps taken towards this direction, its finding doesn't look near (Simmonds & MacLennan, 2005). Several studies have focused on the classification of schools to species taking into account the multifrequency response of the different species (Woillez *et al.*, 2012) and/or traits of school structure and behaviour. At different areas various methodologies such as multivariate approaches (e.g. Haralabous & Georga-

karakos, 1996; Scalabrin *et al.*, 1996; Simmonds *et al.*, 1996; Lawson *et al.*, 2001; Fernandes, 2009) and neural networks (e.g. Haralabous & Georgakarakos, 1996; Simmonds *et al.*, 1996; Robotham *et al.*, 2010) have been successfully applied locally for the classification of schools into species. Nevertheless, the differences in species composition among ecosystems and the high spatiotemporal variability in school characteristics even within a species (Massé *et al.*, 1996; Scalabrin *et al.*, 1996; Iglesias *et al.*, 2003), limit the ability to generalize and apply results in regions other than the study area(s).

Concerning the Mediterranean Sea, few studies have focused on the school structure in relation to the different species observed. Haralabous & Georgakarakos (1996) successfully discriminated anchovy *Engraulis encrasicolus,* sardine *Sardina pilchardus* and horse mackerel *Trachurus trachurus*, in a small area of the North Aegean Sea during 1992 and 1993 by taking into account energetic, bathymetric and morphometric parameters. Similarly, in the Central Mediterranean, D'Elia et al. (2014) were able to classify the same three species as well as a group defined as a mixture of other small pelagics species based on bathymetric and morphometric school descriptors complemented by multifrequency responses. On the other hand in the Spanish Mediterranean waters, high annual variability in school structure was observed but this was more related to the size of the individuals than to species composition (Iglesias *et al.*, 2003).

In the Aegean Sea (Eastern Mediterranean), several acoustic surveys mainly targeting the biomass estimation of anchovy and sardine have taken place since the late '80s. The obtained information has been used as input for stock assessment (e.g., Giannoulaki *et al.*, 2014), food-web (Tsagarakis *et al.*, 2010) and species distribution (e.g., Tugores *et al.*, 2011; Giannoulaki *et al.*, 2013) models as well as for the study of the spatial structure (e.g., Giannoulaki *et al.*, 2006) of these two species. Furthermore, data obtained in the framework of these surveys have been used to investigate aspects of the schooling behaviour, i.e. diel vertical migration of sardine (Giannoulaki *et al.*, 1999) and anchovy (Tsagarakis *et al.*, 2012a) as well as ontogenetic changes in the schooling of sardine (Tsagarakis *et al.*, 2012b).

In the present work we examined differences in the school structure of five small pelagic fish species using data from acoustic surveys (2004-2008) in the North Aegean Sea. Towards the direction of upgrading biomass estimates in the area and adapting fisheries acoustics to the increased needs of ecosystem based management (Trenkel *et al.*, 2011), we applied a classification approach, based on validation trawl samples along with bathymetric, morphometric and energetic schools characteristics (supervised learning approach; Woillez *et al.*, 2012). Our aims were to (a) discriminate schools according to species with a straightforward methodology that would allow us the direct application to the analysis of acoustic data and (b) identify

which school characteristics largely differentiate among species, thus being responsible for their discrimination.
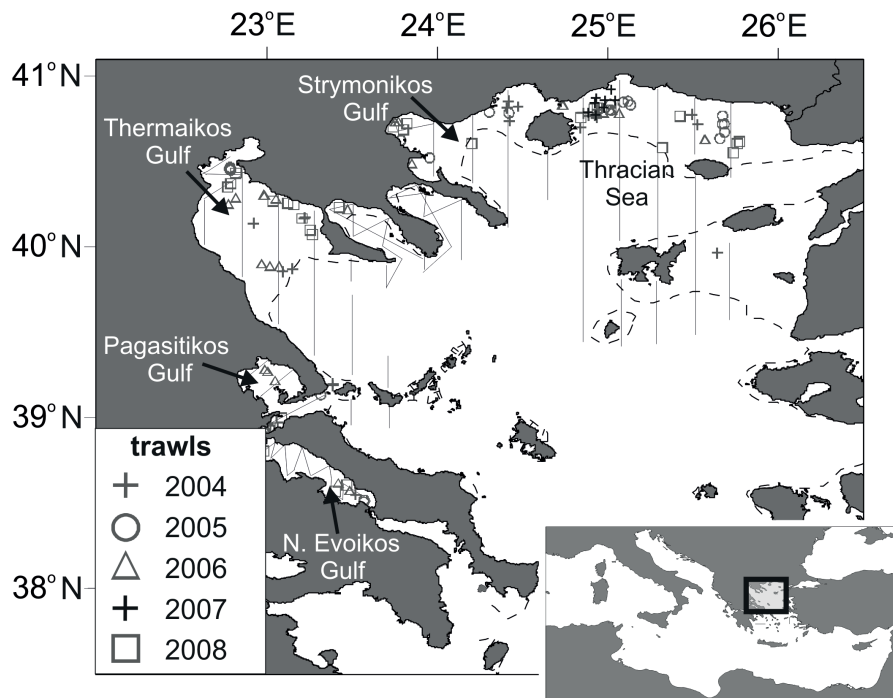
Materials and Methods

Sampling and Acoustic data

Acoustic data were collected on board R/V Philia during five research cruises held during summer within the period 2004-2008 in the North Aegean Sea, Eastern Mediterranean (Fig. 1). Acoustic sampling was performed by means of a scientific split Beam Biosonics DTX echosounder operating at 38 kHz and calibrated following standard techniques (Foote *et al.*, 1987). Acoustic data were recorded at a constant speed of 7-8 nmi h^{-1}. Minimum sampling depth varied between 10 to 20 m depending on the area. Echosounding took place along predetermined transects, roughly perpendicular to the bathymetry and with 10 nautical miles distance to each other (Fig. 1). In enclosed gulfs a random design (zig-zag form) was applied (Fig. 1). In 2007 only a small region in the Thracian Sea was covered while a denser transect design with 5 nautical miles distance was additionally applied in the Thracian Sea in 2008. A grid of CTD stations (10 nmi distant to each other) was sampled concurrently with acoustics. At each station vertical profiles of temperature and salinity were obtained using a SBE-25 internally recording CTD unit (Sea Bird Electronics) and processed using Seasoft-Win32 software. Temperature and salinity values obtained were used for the estimation of echo propagation parameters (sound speed and absorption coefficient).

Pelagic trawl sampling was carried out in parallel and concurrently to the acoustic sampling (Simmonds & MacLennan, 2005) in regions with high fish concentration (Fig. 1). During trawling, vessel speed approximated 4 knots. Trawl catches were sorted to the species level and the total biomass and abundance per species were measured or estimated based on a representative sample when catches were too high. Trawling took place either during daytime so, the recorded schools were assigned to the species caught, or during the night when schools are dispersed, thus catches were used to determine the species composition and the relative abundance in the region (Machias *et al.*, 2013).

The minimum threshold for acquisition was -70dB and acoustic data analysis was performed using Myriax Echoview software. In the echograms the seabed was detected using the software's dedicated algorithm and the same was done for school detection after setting 5 m as minimum length and 1 m as minimum height for candidate schools.

Maximum ping rate is automatically limited by the Biosonics software depending on the echosounded vertical range in order to avoid the effect of shadow bottoms in the acoustic data (Simmonds & MacLennan, 2005). Thus, during echosounding ping rate was set to maximum achievable; indicatively, it was 3.65 pings per second (pps) at 50 m bottom depth, 2.43 pps at 80 m bottom depth and 1.35 pps at 150 m bottom depth. Pulse duration

Fig. 1: Survey design. Symbols indicate the positions of the research trawls in each year. The 100 m isobath is indicated with a dashed line.

was set to 0.4 ms, which allowed us to sufficiently separate two targets at 0.3 m distance apart. Thus only echoes detected in the water layer between 0.5 m above the seabed and up to 5 m below the surface were analyzed. Allocation of schools into species was based on the trawl catches, taking also into account previously gained experience (Simmonds & MacLennan, 2005; Tsagarakis *et al.*, 2012a; Tsagarakis *et al.*, 2012b). Following the analy-

sis of the echograms, a series of positional, energetic and morphometric school descriptors were extracted from the software or estimated as described in Table 1. The basic morphometric descriptors (*Height*, *Length*, *Area* and *Perimeter*) were corrected according to Diner's (2001) algorithm. Schools too small in size to be corrected with Diner's (2001) algorithm represented approximately 10% of our dataser and were discarded from the analysis.

Table 1. Positional (A), energetic (B) and morphometric (C) school descriptors used in the analysis.

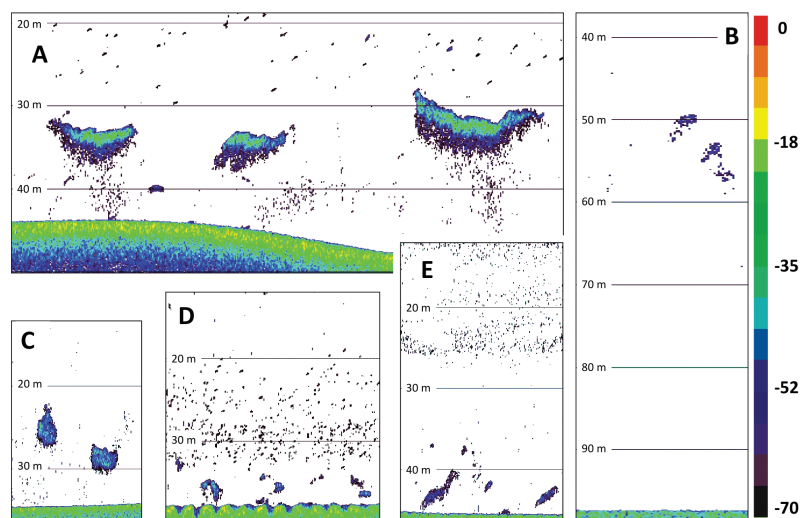| Descriptor	Details
A. Positional	
School Depth	Mean depth of the school (m)
Altitude	Distance between the lower extend of the school and the bottom (m)
B. Energetic	
MVBS	Mean Volume Backscattering Strength (dB re m^2*m^{-3})
C. Morphometric	
Height (H)	Maximum height of the school (m)
Length (L)	Maximum length of the school (m)
Area (A)	2D surface of the school (m^2)
Perimeter (P)	2D perimeter of the school, (m)
Elongation (E)	Length on height ratio (E= L/H)
Fractal Dimension (FD)	Expresses the roughness of school shape. More regular shapes have values closer to one (FD= 2ln*(P/4)/lnA)
Circularity	Expresses the deviation from the circle (C= P^2/4*π*A)
Rectangularity	Expresses the deviation from the rectangular (R=L*H/A)

Data analysis

In order to minimize errors due to possible false assignment of schools into species, we limited our analysis to those schools recorded (i) concurrently to trawling, in cases that the relative abundance of a species in the trawl was higher than 75%, and (ii) along transects in regions where the relative abundance of a species was higher than 75% among small pelagic fish (as verified by experimental fishing in the region). Only schools detected during daytime were used, i.e. we excluded schools recorded during the transitional periods (dawn and dusk) in which school structure has been found to vary significantly (Fréon *et al.*, 1996; Giannoulaki *et al.*, 1999; Tsagarakis *et al.*, 2012a). The number of schools per species that fulfilled the above mentioned criteria is listed in Table 2 for the five research cruises. The majority of the schools concerned anchovy and sardine, however a small number of schools of round sardinella *Sardinella aurita*, blue whiting *Micromessistius poutassou,* and Atlantic mackerel *Scomber scombrus,* was also identified. Examples of schools as seen in echograms are illustrated for these five species in Figure 2.

We applied Discriminant Function Analysis (DFA) to the series of positional, energetic and morphometric school descriptors as described in Table 1. Spearman correlation among the school descriptors, and additionally including bottom depth, was explored in order to i) test for multi-colinearity and ii) exclude the possibility that the observed variability in school descriptors is largely attributed to their correlation with bottom depth. No high multicolinearity ($r > 0.8$) was observed among the parameters with few exceptions (e.g., among *Length*, *Height* and *Area;* Fig. 3), thus no important constrains in the explanatory power of the analysis are expected.

The discriminant analysis was applied separately to data from each cruise as well as to pooled data from all cruises. A forward stepwise selection was applied using the school descriptors as explanatory variables to reveal which school characteristics are included in the final discriminant and classification functions. The significance of the DFA was based on the Wilks' λ and the relative F-statistic (McLachlan, 2004). The percentage of correct classifications of schools into species was assumed as

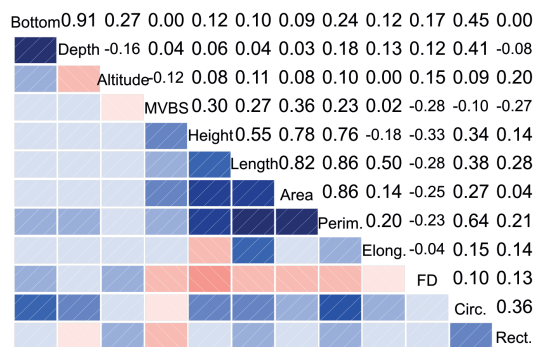**Table 2.** Number of trawls performed in each cruise and number of schools per species included in the analyses.

Cruise	number of trawls	Anchovy	Sardine	Round sardinella	Blue whiting	Mackerel	Total
				Schools per species			
2004	23	117	72				189
2005	27	57	7	3		15	82
2006	37	73	17		18		108
2007	15	120	81				201
2008	40	48	97	16			161
Total	**142**	**415**	**274**	**19**	**18**	**15**	**741**

Fig. 2: Example echograms representing schools of the five species included in the analysis. A: Round sardinella; B: Blue whiting; C: Sardine; D: Anchovy; E: Mackerel. The vertical scale is the same in all echograms but the horizontal one may differ. The colour scale indicates backscattering strength (dB).

Bottom 0.91 0.27 0.00 0.12 0.10 0.09 0.24 0.12 0.17 0.45 0.00
Depth -0.16 0.04 0.06 0.04 0.03 0.18 0.13 0.12 0.41 -0.08
Altitude -0.12 0.08 0.11 0.08 0.10 0.00 0.15 0.09 0.20
MVBS 0.30 0.27 0.36 0.23 0.02 -0.28 -0.10 -0.27
Height 0.55 0.78 0.76 -0.18 -0.33 0.34 0.14
Length 0.82 0.86 0.50 -0.28 0.38 0.28
Area 0.86 0.14 -0.25 0.27 0.04
Perim. 0.20 -0.23 0.64 0.21
Elong. -0.04 0.15 0.14
FD 0.10 0.13
Circ. 0.36
Rect.

Fig. 3: Correlation matrix of the parameters included in the DFA, additionally including *Bottom Depth (Bottom)*. Numbers are correlation coefficients; blue and red hues indicate positive and negative correlations respectively; the darker the colour the higher the correlation; Depth = *School Depth;* Perim. = *Perimeter*; Elong. = *Elongation*; FD = *Fractal Dimension*; Circ. = *Circularity*; Rect. = *Rectangularity.*

an indirect measure of the discriminating power of the DFA. In addition, since the number of correctly classified schools may not reflect the fraction of echo that is well classified, we estimated the backscattering energy (Nautical Area Scattering Coefficient; NASC) corresponding to those schools that were correctly classified by the DFA. This can be considered as an additional indication of the reliability of the methodology for biomass estimations.

The standardized coefficients of the discriminant functions show the relative contribution of each parameter in the discrimination of the groups (McLachlan, 2004). The coefficients of the classification function(s) can be used for the classification of schools that have not been assigned into species as well as for the cross-validation of the results. We used 80% of the schools for each DFA applied while the remaining 20% was used for cross-validation (Lawson *et al.*, 2001). In cases that the number of schools for some species was too low to exclude 20% from the analysis (i.e. in 2005, 2006, 2008 and in the case of pooled data where a limited number of schools of round sardinella, blue whiting and/or mackerel were included in the analysis), cross validation was based on an automated procedure of continuous DFAs

where one observation was not taken into account each time (leave-one-out method) (McLachlan, 2004). The Statgraphics Centurion XV statistical software was used for the stepwise DFA while S-Plus statistical package was used for the leave-one-out cross validation method.

For each species Box-and-Whisker plots were used to graphically represent the values of the school descriptors that were selected during the stepwise procedure of the DFA of pooled data. Significant differences among species for each descriptor were explored with the Kruskal-Wallis test. Finally, Welch t-tests for unequal variances (Welch, 1947) were used to explore pair wise differences between anchovy and sardine, i.e. the two most abundant species.

Results

Results reveal that the DFA based on positional, energetic and morphometric school descriptors can be successfully used to discriminate the schools of the five species that were identified during summer in the North Aegean Sea. Approximately 72% of the schools were correctly classified in the analysis of pooled data while the backscattering echo corresponding to these schools was even higher (>81%) (Table 3). The percentage of correct classifications in terms of number of schools but also in terms of backscattering echo was higher in the analysis per cruise, exceeding 90% in 2004 and 2008 (Table 3). As generally expected, the percentage was also higher when the prior probability of correct assignment was also high, i.e. in cruises that only few species were included in the analysis.

The percentages of correct classifications per species are presented in Table 4 for the analysis of pooled data. Despite the fact that three out of five species were represented only by a small number of schools (i.e. ≤19) in a small number of surveys, DFA correctly classified 94% of blue whiting schools and around 70% of the remaining species (Table 4). These percentages were higher in the analyses of each cruise separately (results for each species are not shown), which was the reason for the overall improved classification in the partial analyses.

In the analysis of pooled data, the first two discriminant functions were capable to discriminate almost 80%

Table 3. DFA results using data from each research cruise and pooled data from all surveys.

Cruise	Wilk's λ	Approximate F	p-value	Prior probability	Number of species	Correct assignments	Cross-validated assignments	Correctly classified backscattering energy
2004	0.213	75.38	<0.001	50%	2	95.36%	89.47%	97.64%
2005	0.142	18.10	<0.001	25%	4	75.61%	75.61%	87.82%
2006	0.127	46.09	<0.001	33%	3	77.78%	76.85%	64.36%
2007	0.559	20.11	<0.001	50%	2	83.75%	78.05%	97.27%
2008	0.098	67.40	<0.001	33%	3	91.93%	90.68%	95.89%
Pooled	0.162	53.58	<0.001	20%	5	71.93%	71.26%	81.39%

Table 4. Classification table of schools per species based on DFA results, using pooled data from all surveys and for all species. Percentages of assignments per species are shown in parentheses. Correct assignments are shown in bold.

Actual group	Classified at group					Group size
	1	2	3	4	5	
1. Anchovy	**296 (71.33%)**	83 (20%)	2 (0.48%)	5 (1.2%)	29 (6.99%)	415
2. Sardine	37 (13.5%)	**197 (71.9%)**	6 (2.19%)	0 (0%)	34 (12.41%)	274
3. Round sardinella	0 (0%)	6 (31.58%)	**13 (68.42%)**	0 (0%)	0 (0%)	19
4. Blue whiting	1 (5.56%)	0 (0%)	0 (0%)	**17 (94.44%)**	0 (0%)	18
5. Mackerel	5 (33.33%)	0 (0%)	0 (0%)	0 (0%)	**10 (66.67%)**	15

Table 5. DFA classification and discriminant function coefficients, using pooled data from all surveys. Values of the two parameters mostly contributing to each discriminant function are shown in bold. DF: Discriminant function, 1: Anchovy, 2: Sardine, 3: Round sardinella, 4: Blue whiting, 5: Mackerel, FD: Fractal Dimension.

Descriptor	Classification function coefficients					DF Standardized Coefficients			
	1	2	3	4	5	DF 1	DF 2	DF 3	DF 4
School Depth	0.352	0.264	0.308	0.523	0.338	0.088	**0.810**	0.587	0.438
Altitude	-0.061	0.026	0.273	0.961	-0.102	**0.890**	0.414	-0.249	-0.138
MVBS	-2.747	-2.643	-2.312	-2.771	-2.822	0.223	-0.332	0.066	-0.284
Area	0.019	0.025	0.170	0.047	0.024	**0.803**	**-0.528**	**0.924**	0.212
Height	2.019	2.040	0.296	1.459	2.135	-0.466	0.151	**-0.599**	-0.024
Circularity	-0.212	-0.211	-0.229	-0.304	-0.268	-0.262	-0.220	0.059	**-0.657**
Rectangularity	1.193	2.036	3.597	1.754	5.564	0.194	-0.262	-0.144	**1.010**
FD	1.946	1.684	1.695	3.905	1.276	0.132	0.281	0.035	-0.142
CONSTANT	-87.064	-80.729	-81.962	-122.555	-99.986				
Relative percentage						45.41%	33.67%	16.74%	4.18%

of the schools (Table 5). Eight parameters were included in the DFA, as the ones contributing to the discrimination of schools according to species. These included bathymetric (*School Depth*, *Altitude*), energetic (*MVBS*) and morphometric school descriptors, the latter expressing both size (*Area*, *Height*) and shape (*Circularity*, *Rectangularity* and *Fractal Dimension*) traits. Due to the low multicolinearity among the parameters, the standardized discriminant function coefficients can be considered that reliably assess the relative importance of the predictor variables. *Altitude*, *School Depth* and *Area* were the parameters mainly contributing, as revealed by the standardized coefficients of the discriminant functions (Table 5). Table 5 also shows the coefficients for each classification function (one per species) derived from the analysis.

The Kruskal-Wallis tests revealed statistically significant differences ($p<0.001$) in all the descriptors included in the DFA, while pair wise comparisons (Welch t-test) between anchovy and sardine also showed significant differences in all the parameters examined (results not shown). The values of the school descriptors revealed some species-specific traits and are presented as Box-and-Whisker plots in Figure 4.
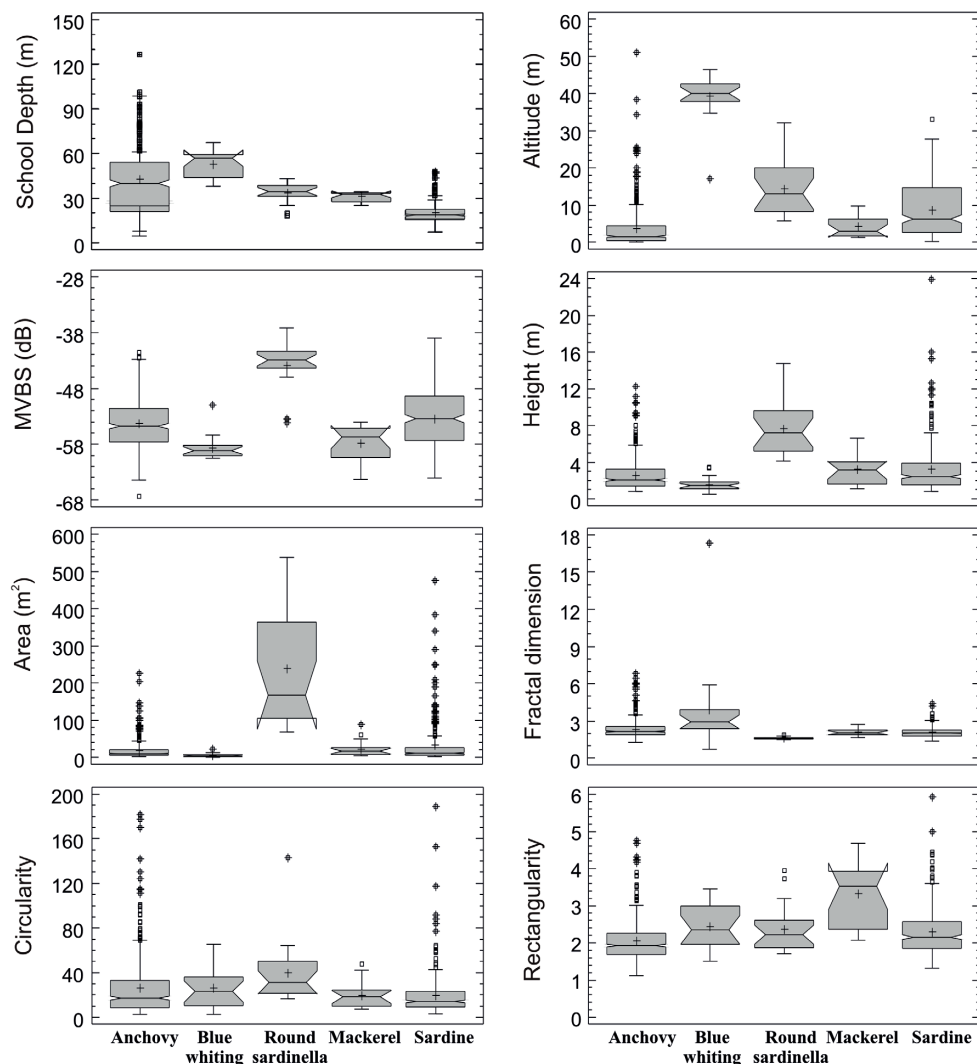
Anchovy schools were found closer to the bottom compared to the other species and exhibited a wide range of bathymetric distribution, exceeding 100 m *School Depth* in several cases. They presented intermediate values in backscattering strength (*MVBS*), size and shape, with the exception of *Rectangularity* since they formed the most rectangular schools (Fig. 4). Sardines were closer to the surface (lower *School Depth*) than all the other species and formed bigger schools, reflecting more energy than anchovy. In addition, they formed the most circular schools.

Schools of blue whiting were located deeper in the water column than the rest of the species, but were also found in higher distance from the bottom (higher *Altitude*; Fig. 4). They reflected the less energy (*MVBS*), they were the smallest (low *Height* and *Area*) and presented the most irregular shape (high *Fractal* Dimension) among the species considered. Schools of round sardinella backscattered the higher energy (*MVBS*) and this was also true for school size (*Height, Area*). As concerns school shape, round sardinella formed the most regular schools (low *Fractal Dimension*) and deviated the most from the circle (high *Circularity*). Finally, schools of mackerel presented intermediate values in all school descriptors with the exception of *Rectangularity* (i.e., they deviated the most from the rectangular shape; Fig. 4).

Discussion

A fish school is an aggregation of constantly moving individuals, responding to external stimuli (e.g., preda-

Fig. 4: Box-and-Whisker plots of school descriptors per species, for pooled data from all surveys. Box range indicates 50% of observations. Notches and crosses inside the boxes indicate medians and means respectively. Vertical lines extend from the lower (or higher) quartile to the smallest (or higher) point within 1.5 interquartile ranges. Points outside the boxes indicate outliers.

tor and prey presence), which results in great variability of its basic structure (Misund, 1993). Nevertheless, some general patterns in the school structure of each fish species can be seen and this is the main principle that the current work is based on. Several studies have focused on school typology (Reid, 2000) and classification of schools to species (e.g., Haralabous & Georgakarakos, 1996; Scalabrin *et al.*, 1996; Simmonds *et al.*, 1996; Lawson *et al.*, 2001; Fernandes, 2009; Korneliussen *et al.*, 2009; Robotham *et al.*, 2010). However, the variability in the community composition among ecosystems along with the high interspecific (Brehmer *et al.*, 2007) and even intraspecific variability (Massé *et al.*, 1996;

Scalabrin *et al.*, 1996) in schooling impair the application of results in regions other than the study area.

Our results show that schools in the North Aegean Sea can be adequately (71.9% in terms of number of schools for pooled data and up to 95% for some cruises) classified to species based on descriptors of the school structure and behaviour. In terms of backscattering echo, these schools correspond to even higher percentages implying that most of the larger and/or denser schools were correctly classified. Several studies have reported percentages of correct classifications ranging from 57% (Scalabrin *et al.*, 1996) to more than 90% (e.g. Haralabous & Georgakarakos, 1996; Simmonds *et al.*, 1996;

Lawson *et al.*, 2001). This percentage largely depends on species-specific behaviour, on the number of species and on the spatiotemporal scale taken into account (i.e. sampling area covered, seasons and years taken into account); the more the spatiotemporal scale is reduced the higher the successful classification of schools. In our study area the percentage of correct classifications increased in the partial analyses of the research cruises (75.6%-95.36% in terms of number of schools) compared to the analysis of pooled data from all surveys (71.9%). Similarly, Scalabrin *et al.* (1996) were able to increase the correct classification of three small pelagic fish schools (anchovy, sardine and horse mackerel) in the Bay of Biscay from 57% to 98% when time and space scales were reduced. Reducing spatiotemporal scales not only restricts the variability in species composition but also the intraspecific variability. Schooling behaviour of a species has been shown to be affected, among others, by the age structure of the populations (e.g. Iglesias *et al.*, 2003; Tsagarakis *et al.*, 2012b), the abiotic environmental conditions (e.g. hydrography: Massé, 1996; substrate type: D'Elia *et al.*, 2009; bottom depth: Tsagarakis *et al.*, 2012a) as well as the presence of other species (Massé, 1996; Massé *et al.*, 1996). Thus, in the case of the separate research cruises, the variability in one or more of these parameters is reduced, resulting in more similar school structures within a given species and in increased discriminating power of the analysis.

The high percentage of correct classifications in each research cruise is also partly explained by the smaller number of species taken into account (2-4 species) as opposed to the analysis of the pooled data (5 species). For example, in 2004 when only anchovy and sardine schools were considered, correct classification exceeded 95%. These are in line with results from the Bay of Biscay; when the analysis was restricted to anchovy and sardine, the discrimination power was improved (Scalabrin *et al.*, 1996). However, in our cruise of 2007 despite (i) the inclusion of only anchovy and sardine in the analysis and (ii) the small spatial scale examined, the percentage of correct classification was lower (83.75%) than in 2004 (when also only anchovy and sardine were considered). Again, the intraspecific variability seems to be the cause; this cruise focused on juvenile sardines which form discrete schools in comparison to the adults (Tsagarakis *et al.*, 2012b). The high presence of schools of both juveniles and adults seems to increase the variability of the group resulting in some false classification of schools to species. Specifically, juvenile sardines form smaller schools than the adult ones (Tsagarakis *et al.*, 2012b) and in this sense they present more similarities with the generally smaller anchovy schools, which reduces the classification power of the DFA.

Percentages of correct classification also vary according to species. In our case this can be partly explained by the relatively low representation of some species in the data. Specifically, the small number of round sardinella and mackerel schools included in the analysis may be responsible for the relatively poor discrimination (<69% in pooled data) of these species; nevertheless, this was not the case for blue whiting which had also low representation in the data. Even though, currently, these species are not primary targets for fisheries (Tsagarakis *et al.*, 2012c) nor for acoustic surveys (MEDIAS, 2008) in the area, improving and upgrading their biomass estimations is important especially under the framework of the Ecosystem Approach to Fisheries (Trenkel *et al.*, 2011). It seems that more data are needed for reliable biomass estimations of these species, thus these limitations should be taken into account in the application of the acoustic sampling as well as in data analysis.

Usually, the analyses of the acoustic data for biomass estimations include either (i) allocation of total echo per sampling unit to the catches of the nearest haul(s), or (ii) assignment of the recorded schools into species based on aspects of school structure taking also into account previously gained experience, or a balanced use of (i) and (ii). Both methods include possible biases due to e.g. different catchability of species by the trawling gear or the subjective assignment of a specific school to a certain species. DFA results, specifically the classification functions, can be directly applied to classify schools that have not been reliably assigned to species after ground-truthing (e.g. in our case, the schools that didn't fulfil the criteria to be included in the analysis). This can lead to the upgrade of biomass estimates since it constitutes an objective, more automated and less time consuming procedure. The cross-validation of the results further supports the strength of the analysis since correct classifications of schools used for cross-validation were at similar levels with the ones included in the analysis. As mentioned above, the analysis was more effective when analyzing data from the separate research cruises. Thus, following the echogram scrutinization and school detection, the analysis can be performed in a set of classified schools and the derived classification functions can be applied to the remaining schools recorded during the entire cruise. In cases that some species are caught in experimental hauls without however being possible to assign to them specific schools, general classification functions derived from analysis of pooled data (where these species are likely to be present) could be applied. This would allow the allocation of schools with high probability to these species. In this case, the area that they were recorded and the respective trawl catches should be taken into account in order to finally decide whether these ambiguous schools should be classified into a specific species or the echo should be directly allocated to the haul catches in the area. The latter could be also applied for schools not possible to be included in the analysis, such as in the case of schools that cannot be corrected according to Diner's (2001) algorithm.

158

Medit. Mar. Sci., 16/1, 2015, 151-161

The effectiveness of the discrimination of schools that belong to different species implies that the interspecific variability is higher than the intraspecific one. Significant differences among species were identified for all the parameters selected during the stepwise procedure of the DFA. Bathymetric descriptors (*School Depth*, *Altitude*) were among the ones contributing the most to the discrimination both in the analysis of the pooled data and in the separate analyses of each research cruise (results not shown) in line with other similar studies (Lawson *et al.*, 2001; Charef *et al.*, 2010; Robotham *et al.*, 2010; D'Elia *et al.*, 2014). *School Depth* and the distance from the bottom (*Altitude*) reveal species-specific behaviours that may be related to micro-habitat preferences (e.g. position in the water column). For example, sardine schools exhibit a different behaviour and are found higher in the water column (lower *School Depth* and higher *Altitude*) in relation to anchovy, as also described by others (Scalabrin *et al.*, 1996; D'Elia *et al.*, 2014). Furthermore, large scale habitat preferences may be revealed for some species; blue whiting schools are found deeper in the water column but also in higher distance from the bottom, which reflect its distribution in areas with deeper waters (Katsanevakis *et al.*, 2009).

Apart from the bathymetric descriptors, energetic (*MVBS*) and morphometric ones, the latter both related to size and shape, contributed to the discrimination of schools. However, they had generally lower relative contribution - especially the ones related to shape - compared to the remaining parameters. This is related to the high plasticity observed in fish schools (Misund, 1993) affected by environmental, behavioural, ontogenetic and physiological traits (e.g. Machias & Tsimenides, 1996; Massé *et al.*, 1996; Nøttestad *et al.*, 1996; Tsagarakis *et al.*, 2012b). Nevertheless some general patterns that remain more stable per species can be identified, e.g. that sardines form schools of greater height which backscatter more energy in relation to anchovy.

The consideration of additional descriptors such as multifrequency differences (Woillez *et al.*, 2012) might improve the discriminating power of the analysis, even if they play a secondary role, as has been already shown for anchovy and sardine schools in the Mediterranean (D'Elia *et al.*, 2014). Furthermore, parameters that may be related to species distribution and habitat selection, such as geographical coordinates and environmental parameters have been used by others in order to improve the classification of schools into species (e.g. Lawson *et al.*, 2001). The use of such variables is limited since (i) they deviate from the school level and they mainly refer to patterns at the species level and (ii) their inclusion in the analysis requires a relatively homogenous spatial distribution of the data. Their use can prove useful in cases that the species considered share similar school traits (Robotham *et al.*, 2010), which is not the case in the N. Aegean Sea as shown in our results.

Finally, our analysis allows for comparisons of species-specific school traits with other areas in the Mediterranean and the Atlantic, at least for the two most studied species, anchovy and sardine. Anchovy schools were mainly found in the water column to depths up to 70 m during daytime, as also reported for the Atlantic (Scalabrin *et al.*, 1996), while distance from the bottom was usually low (below 10 m) but in several cases it exceeded 20 m in the North Aegean Sea, Sicily straits (D'Elia *et al.*, 2009) and the Bay of Biscay (Massé, 1996). Moreover, according to our findings, anchovy schools were smaller in the North Aegean Sea than in other areas (Massé *et al.*, 1996, D'Elia *et al.*, 2009) and reflected less energy compared to the Atlantic (Scalabrin *et al.*, 1996). *Elongation* was similar in all areas however in the North Aegean Sea, school shape was the least regular (Massé *et al.*, 1996, Scalabrin *et al.*, 1996, D'Elia *et al.*, 2009). As concerns sardine, schools were found high in the water column (10-50 m *School Depth*) and in relatively high *Altitude* both in the North Aegean and the Atlantic (Scalabrin *et al.*, 1996; Muino *et al.*, 2003); this reflects its distribution in shallow areas during summer (Tugores *et al.*, 2011). Differences are observed in the backscattering echo with *MVBS* being lower in the North Aegean Sea (-54.5 dB) compared to the Atlantic (>-40 dB; Scalabrin *et al.*, 1996; Zwolinski *et al.*, 2007). School size and shape in our study was in the range of the values already reported for the Atlantic (Scalabrin *et al.*, 1996, Zwolinski *et al.*, 2007). The aforementioned differences imply a high plasticity in bathymetric, energetic and morphometric schools traits within a species. Part of this variability is explained by (i) genetic differences in anchovy populations in the Aegean Sea, the Central Mediterranean and the Atlantic (Magoulas *et al.*, 2006) and (ii) the existence of at least two discrete sub-species of sardine in the Mediterranean and the Atlantic (Gonzalez & Zardoya, 2007). Intraspecific differences in biology (e.g. growth and maturity of sardine populations; Silva, 2003; Silva *et al.*, 2006) and possibly in behaviour may also affect school formation. Still, even in a single study area the intraspecific variability is high, as evidenced by the large range of values of some school descriptors in the North Aegean Sea, and this can affect the effectiveness of the school classification. The exploration of this variability in relation to environmental parameters and behavioural patterns is important and could further improve sampling design and biomass estimates.

References

Brehmer, P., Gerlotto, F., Laurent, C., Cotel, P., Achury, A. *et al.,* 2007. Schooling behaviour of small pelagic fish: phenotypic expression of independent stimuli. *Marine Ecology Progress Series*, 334, 263-272.

Charef, A., Ohshimo, S., Aoki, I., Al Absi, N., 2010. Classification of fish schools based on evaluation of acoustic descriptor characteristics. *Fisheries Science*, 76 (1), 1-11.

D'Elia, M., Patti, B., Bonanno, A., Fontana, I., Giacalone, G. *et al.,* 2014. Analysis of backscatter properties and application of classification procedures for the identification of small pelagic fish species in the Central Mediterranean. *Fisheries Research*, 149, 33–42.

D'Elia, M., Patti, B., Sulli, A., Tranchida, G., Bonanno, A. *et al.,* 2009. Distribution and spatial structure of pelagic fish schools in relation to the nature of the seabed in the Sicily Straits (Central Mediterranean). *Marine Ecology*, 30 (Suppl. 1), 151–160.

Diner, N. 2001. Correction on school geometry and density: approach based on acoustic image simulation. *Aquatic Living Resources*, 14, 211−222.

Fernandes, P.G., 2009. Classification trees for species identification of fish-school echotraces. *ICES Journal of Marine Science*, 66 (6), 1073-1080.

Foote, K., Knudsen, H., Vestnes, G., Maclennan, D., Simmonds, E., 1987. *Calibration of acoustic instruments for fish density estimation: a practical guide,* International Council for the Exploration of the Sea.

Fréon, P., Gerlotto, F., Soria, M., 1996. Diel variability of school structure with special reference to transition periods. *ICES Journal of Marine Science*, 53, 459-464.

Fréon, P., Misund, O.E., 1999. *Dynamics of Pelagic Fish Distribution and Behaviour: Effects on Fisheries and Stock Assessment,* Blackwell Science Ltd, Oxford, 360 pp.

Garcia, S. M., Zerbi, A., Aliaume, C., Do Chi, T., Lasserre, G., 2003. The ecosystem approach to fisheries. Issues, terminology, principles, institutional foundations, implementation and outlook. *In* FAO Fisheries Technical Paper No. 443, p. 71 FAO, Rome.

Giannoulaki, M., Ibaibarriaga, L., Antonakakis, K., Uriarte, A., Machias, A. *et al*., 2014. Applying a two-stage Bayesian dynamic model to a short lived species, the anchovy, in the Aegean Sea (Eastern Mediterranean). Comparison with an Integrated Catch at Age stock assessment model. *Mediterranean Marine Science*, 15, 350-365.

Giannoulaki, M., Iglesias, M., Tugores, P., Bonanno, A., Patti, B. *et al*., 2013. Characterising the potential habitat of European anchovy Engraulis encrasicolus in the Mediterranean Sea, at different life stages. *Fisheries Oceanography*, 22, 69-89.

Giannoulaki, M., Machias, A., Koutsikopoulos, C., Somarakis, S., 2006. The effect of coastal topography on the spatial structure of anchovy and sardine. *ICES Journal of Marine Science*, 63, 650-662.

Giannoulaki, M., Machias, A., Tsimenides, N., 1999. Ambient luminance and vertical migration of the sardine *Sardina pilchardus*. *Marine Ecology Progress Series*, 178, 29-38.

Gonzalez, E., Zardoya, R., 2007. Relative role of life-history traits and historical factors in shaping genetic population structure of sardines (*Sardina pilchardus*). *BMC Evolutionary Biology*, 7 (1), 197.

Haralabous, J., Georgakarakos, S., 1996. Artificial neural networks as a tool for species identification of fish schools. *ICES Journal of Marine Science*, 53 (2), 173-180.

Horne, J.K., 2000. Acoustic approaches to remote species identification: a review. *Fisheries Oceanography*, 9 (4), 356-371.

Iglesias, M., Carrera, P., Muino, R., 2003. Spatio-temporal patterns and morphological characterisation of multispecies pelagic fish schools in the North-Western Mediterranean Sea. *Aquatic Living Resources*, 16 (6), 541-548.

Katsanevakis, S., Maravelias, C.D., Damalas, D., Karageorgis, A. P., Tsitsika, E.V. *et al*., 2009. Spatiotemporal distribution and habitat use of commercial demersal species in the eastern Mediterranean Sea. *Fisheries Oceanography*, 18 (6), 439-457.

Korneliussen, R.J., Heggelund, Y., Eliassen, I.K., Johansen, G.O., 2009. Acoustic species identification of schooling fish. *ICES Journal of Marine Science*, 66 (6), 1111-1118.

Lawson, G.L., Barange, M., Fréon, P., 2001. Species identification of pelagic fish schools on the South African continental shelf using acoustic descriptors and ancillary information. *ICES Journal of Marine Science*, 58, 275-287.

Machias, A., Pyrounaki, M.M., Leonori, I., Basilone, G., Iglesias, M. *et al*., 2013. Catch of pelagic hauls in Mediterranean acoustic surveys: Is it the same between day and night? *Scientia Marina*, 77, 69-77.

Machias, A., Tsimenides, N., 1996. Anatomical and physiological factors affecting the swim-bladder cross-section of the sardine *Sardina pilchardus*. *Canadian Journal of Fisheries and Aquatic Sciences*, 53, 280-287.

Magoulas, A., Castilho, R., Caetano, S., Marcato, S., Patarnello, T., 2006. Mitochondrial DNA reveals a mosaic pattern of phylogeographical structure in Atlantic and Mediterranean populations of anchovy (Engraulis encrasicolus). *Molecular Phylogenetics and Evolution*, 39 (3), 734-746.

Massé, J., 1996. Acoustic observations in the Bay of Biscay: Schooling, vertical distribution, species assemblages and behaviour. *Scientia Marina*, 60 (Suppl. 2), 227-234.

Massé, J., Koutsikopoulos, C., Patty, W., 1996. The structure and spatial distribution of pelagic fish schools in multispecies clusters: an acoustic study. *ICES Journal of Marine Science*, 53, 155-160.

McLachlan, G.J., 2004. *Discriminant Analysis and Statistical Pattern Recognition,* John Wiley & Sons, New York, 526 pp.

MEDIAS, 2008. *Adoption of a common protocol for MEDiterranean Acoustic Surveys (MEDIAS). Steering Committee Report.* no, 15 pp.

Misund, O.A., 1993. Dynamics of moving masses: variability in packing density, shape, and size among herring, sprat, and saithe schools. *ICES Journal of Marine Science*, 50 (2), 145-160.

Muino, R., Carrera, P., Iglesias, M., 2003. The characterization of sardine (Sardina pilchardus Walbaum) schools off the Spanish-Atlantic coast. *ICES Journal of Marine Science*, 60 (6), 1361-1372.

Nøttestad, L., Aksland, M., Beltested, A., Ferno, A., Johannessen, A. *et al*., 1996. Schooling dynamics of Norwegian spring spawning herring (*Clupea harengus* L.) in a coastal spawning area. *Sarsia*, 80, 277-284.

Plagányi, É.E. 2007. *Models for an ecosystem approach to fisheries,* FAO, 477, Rome 108 pp.

Reid, D.G., 2000. *Report on Echo Trace Classification*. ICES Cooperative Report no 238, 106 pp.

Robotham, H., Bosch, P., Gutiérrez-Estrada, J.C., Castillo, J., Pulido-Calvo, I., 2010. Acoustic identification of small pelagic

fish species in Chile using support vector machines and neural networks. *Fisheries Research*, 102 (1-2), 115-122.

Scalabrin, C., Diner, N., Weill, A., Hillion, A., Mouchot, M.-C. 1996. Narrowband acoustic identification of monospecific fish shoals. *ICES Journal of Marine Science*, 53(2), 181-188.

SGMED, 2009. *Report of the SGMED-09-02 Working Group on the Mediterranean - Part I*. E. Commission, no, 846 pp.

Silva, A. 2003. Morphometric variation among sardine (*Sardina pilchardus*) populations from the northeastern Atlantic and the western Mediterranean. *ICES Journal of Marine Science*, 60, 1352-1360.

Silva, A., Santos, M.B., Caneco, B., Pestana, G., Porteiro, C. *et al*., 2006. Temporal and geographic variability of sardine maturity at length in the northeastern Atlantic and the western Mediterranean. *ICES Journal of Marine Science*, 63 (4), 663-676.

Simmonds, E.J., Armstrong, F., Copland, P.J., 1996. Species identification using wideband backscatter with neural network and discriminant analysis. *ICES Journal of Marine Science*, 53, 189-195.

Simmonds, E.J., MacLennan, D.N., 2005. *Fisheries acoustics: theory and practice,* Blackwell Publishing, 437 pp.

Trenkel, V.M., Ressler, P.H., Jech, M., Giannoulaki, M., Taylor, C., 2011. Underwater acoustics for ecosystem-based management: state of the science and proposals for ecosystem indicators. *Marine Ecology Progress Series*, 442, 285-301.

Tsagarakis, K., Coll, M., Giannoulaki, M., Somarakis, S., Papaconstantinou, C. *et al*., 2010. Food-web traits of the North Aegean Sea ecosystem (Eastern Mediterranean) and comparison with other Mediterranean ecosystems. *Estuarine, Coastal and Shelf Science*, 88, 233-248.

Tsagarakis, K., Giannoulaki, M., Somarakis, S., Machias, A., 2012a. Variability in positional, energetic and morphometric descriptors of European anchovy Engraulis encrasicolus schools related to patterns of diurnal vertical migration. *Marine Ecology Progress Series*, 446, 243-258.

Tsagarakis, K., Pyrounaki, M.M., Giannoulaki, M., Somarakis, S., Machias, A., 2012b. Ontogenetic shift in the schooling behaviour of sardines, Sardina pilchardus. *Animal Behaviour*, 84, 437-443.

Tsagarakis, K., Vassilopoulou, V., Kallianiotis, A., Machias, A., 2012c. Discards of the purse seine fishery targeting small pelagic fish in the Eastern Mediterranean Sea. *Scientia Marina*, 76 (3), 561-572.

Tugores, M. P., Giannoulaki, M., Iglesias, M., Bonanno, A., Ticina, V. *et al*., 2011. Habitat suitability modelling for sardine *Sardina pilchardus* in a highly diverse ecosystem: the Mediterranean Sea. *Marine Ecology Progress Series*, 443, 181-205.

Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E. *et al*., 2003. Remote sensing for biodiversity science and conservation. *Trends in Ecology & Evolution*, 18 (6), 306-314.

Valavanis, V., Pierce, G., Zuur, A., Palialexis, A., Saveliev, A. *et al*., 2008. Modelling of essential fish habitat based on remote sensing, spatial analysis and GIS. *Hydrobiologia*, 612 (1), 5-20.

Welch, B.L., 1947. The Generalization of 'Student's' Problem when Several Different Population Variances are Involved. *Biometrika*, 34 (1-2), 28-35.

Woillez, M., Ressler, P.H., Wilson, C.D., Horne, J.K., 2012. Multifrequency species classification of acoustic-trawl survey data using semi-supervised learning with class discovery. *The Journal of the Acoustical Society of America*, 131 (2), EL184-EL190.

Zwolinski, J., Morais, A., Marques, V., Stratoudakis, Y., Fernandes, P.G., 2007. Diel variation in the vertical distribution and schooling behaviour of sardine (Sardina pilchardus) off Portugal. *ICES Journal of Marine Science*, 64 (5), 963-972.