International Council for the Exploration of the Sea

FISH CAPTURE COMMITTEE
C.M. 1993/B:9

FISH-SCHOOL SPECIES IDENTIFICATION USING A NEURAL NETWORK

by

Haralabous J. and Georgakarakos S.

Institute of Marine Biology of Crete (IMBC)
Department of Hydroacoustics
P.O.Box 2214; GR-71003 Heraklion Crete, Greece

SUMMARY

Fish-schools of Sardines, Anchovies and Horse-Mackerel can be discriminated from each other, using processed data from hydroacoustic surveys. Back-propagation artificial neural networks can be trained to classify such schools reliably, even in the presence of significant overlaps in the characteristics of schools.

1. INTRODUCTION

The development of fish-school identification techniques based on hydroacoustic information, is firmly connected with the reduction of error in biomass estimation. Echogram scrutinising methods, based on the concurrent trawling data and human experiences are time consuming and subjective. Most of the recent improvements attempt to extract from the backscattered echo signals, a set of quantitative parameters, that could describe sufficiently the structure of particular fish aggregations (Diner et al 1989, Georgakarakos et al 1993) or "acoustic populations" (Gerlotto and Freon 1988). Such approaches improve the objectivity in estimations, reduce the consuming of time, and can also provide a base for predictions. If the construction of schools under certain conditions, could be considered species-identical, then it would be possible to predict the species' identity from the associated descriptors.

The selection of the best descriptors, and the accuracy of classification predictions, are the two main implicated problems, correlated to each other. Classical statistical procedures, such as principal components' analysis (PCA) and especially, discriminant function analysis (DFA), are the most common performed techniques in this area (Scalabrin et al 1991). Besides, the strict prerequisites (multivariate normality of distributions, equality of the covariance matrices, etc.) eliminate the reliable use of the above procedures in many circumstances.

The use of artificial neural networks (ANN) does not demand any assumptions on the kind of distributions and is a rather new technique in fish-school identification and classification problems. The aim of the present study is to develop a neural network that can generate the appropriate associations between different school parameters and species' identity, so as to be possible for reliable predictions, and finally to discuss the contribution of this method to the classification of small pelagic fish.

2. METHODS

2.1 Data Acquisition

Our school data obtained from hydroacoustic surveys in Thermaikos Gulf in 1991 and 1992. Data collection at sea was performed using Biosonics dual beam equipment operated at 120 kHz. They have been analysed by using "SCHOOL" software, developed in IMBC (Georgakarakos and Paterakis 1993), in order to identify school formations and to extract the required parameters. Up to now 3420 schools have been encountered, digitised and analysed with this software. More than 90% of these schools belong to the three most common species in this area:

a. *Sardina pilchardus* (Sardine),
b. *Engraulis encrasicolus (Anchovy)*,
c. *Trachurus mediterraneus* (Horse mackerel).

Besides, only 270 of the above schools (8%) were identified with the highest degree of certainty. We have chosen those schools detected during trawling, potentially caught by the trawl, and when the catch was monospecific. More than 30 parameters could be calculated by "SCHOOL" software classified into three groups: morphological, energetic and spatio-temporal (Table 1).

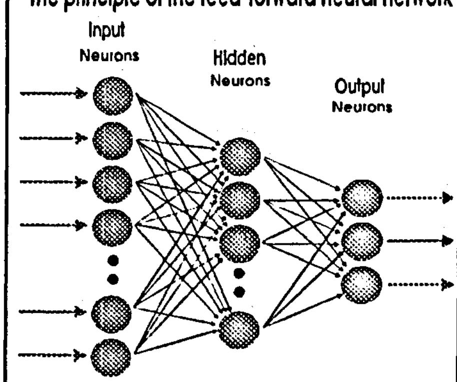
	Descriptors	Full name	Units
1	ID	School ID	-
2	ELEM	Number of pixels/school	-
3	H	Height of school	m
4	DMEAN	mean Depth of school	m
5	PINGS	number of pings/school	-
6	BOT	mean Bottom Depth	m
7	L	Length of school	m
8	ELON	Elongation of school	-
9	AREA	Area of school	m^2
10	SVST	Total SVS of school	V^2
11	SVS	mean SVS of school	V^2
12	SMAX	maximum SVS of school	V^2
13	SSD	Standard Deviation of SVS	-
14	SCV	Coeff. of variation of SVS	-
15	CROWD	Mean Crowding of school	-
16	PATCH	Patchiness of school	-
17	IOD	Index of Dispersion	-
18	K	Clumping Coefficient	-
19	AMIN	minimum Altitude	m
20	AMAX	maximum Altitude	m

Table 1: *The main parameters given by the "SCHOOL" program*

2.2 Neural Network Structure

The artificial neural network (ANN) uses a highly interconnected group of simulated neurons that process information in parallel. The main concept of an ANN is to learn from experience (not from programming) by creating its own internal representations of reality based on raw information given to it (Lawrence 1993). The basic functions of an ANN are: training, testing, and predicting.

Figure 1:
The principle of the feed-forward neural network



SCHOOLBRAIN developed in IMBC, is a *back-propagation* supervised neural network application with a *sigmoid* transfer function. It was developed by the use of a commercial neural network simulator "BrainMaker Professional™ version 2.5" (California Scientific Software©) on the IBM-PCs environment. It has three *layers*:

i. the *input* layer, with all the descriptors of a school (except the species' ID) as input neurons,
ii. the *hidden* layer, with a variant number of neurons, and
iii. the *output* layer, with three neurons representing the three different species.

Neurons in a given layer do not connect to each other. According to the feed-forword concept, the neurons take their inputs from the previous layer only, and send outputs only to the next layer (Figure 1). For this reason a feed-forword network can compute a result very quickly. Back propagation algorithm makes the network learn by correcting the connections, based on the error at the output. Correction signals propagate back through network during training. As training progresses, the amount of error is minimized.

During training SCHOOLBRAIN takes as input every case of school separately (one *training fact* at a time) and gives an *actual output pattern* which is the prediction of species' ID. Before taking the next fact, it compares this output with the *desired (known) output pattern*. If there is a difference between these two patterns (*bad output*), the *weights* are changed to reduce the difference. The amount of the change to the weights is estimated by the *Delta Rule* (see Lawrence 1993). Reading cases step by step and comparing the actual output with the desired one, the network becomes more precise after a number of rounds.

SCHOOLBRAIN can be modified by the user to meet different sets of learning parameters like: training tolerance, learning rate, noise thresholds, smoothing, ability to add hidden neurons if necessary during training, etc. This is very useful for experimental purposes.

3. RESULTS

3.1. Training

The number of the output neurons is the most critical point in training. If we use only one neuron, assigning the different species to different values of the same neuron, the training process becomes difficult and the predictability very poor. The best solution is given by three different neurons, for every different species' ID.

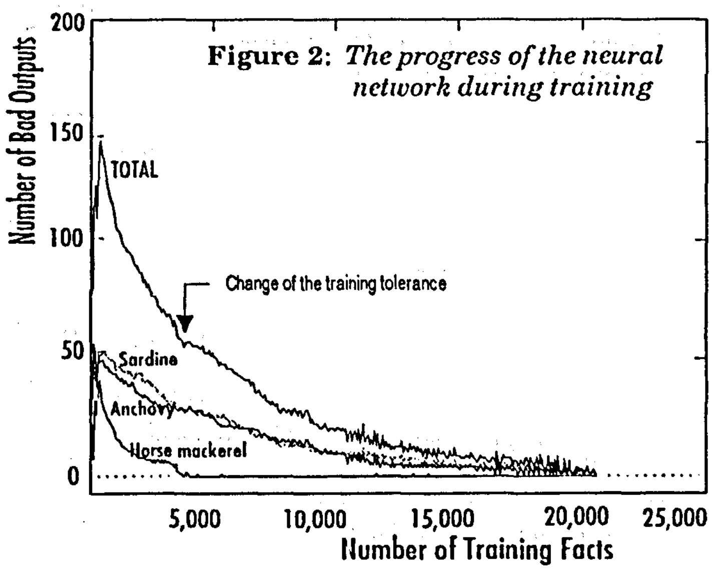


Figure 2: *The progress of the neural network during training*

Experiments using different subsets of the available data, showed that the number of training rounds became higher as the amount of data was enlarged, although the amount of training cases is positively connected with the accuracy of predictions.

In all experiments we discovered that SCHOOLBRAIN could easily be trained to discriminate schools of Horse mackerels from other schools. Discriminating between Sardine and Anchovy schools needed almost five times more training facts (Figure 2).

3.2. Testing and Predicting

Reserving 5-10% of the data from the training patterns, is the best way to test the network. We can not use for testing purposes, data already used for training. Depending on other learning settings, the testing predictions were good for 65 to 99% of the testing cases. Our experiments indicated that best testing predictions came from representative samples of cases.

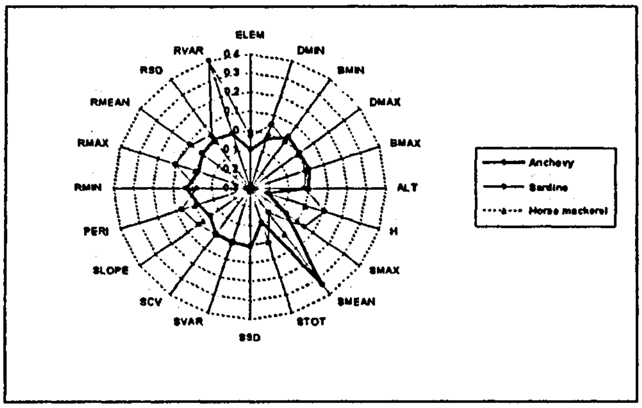


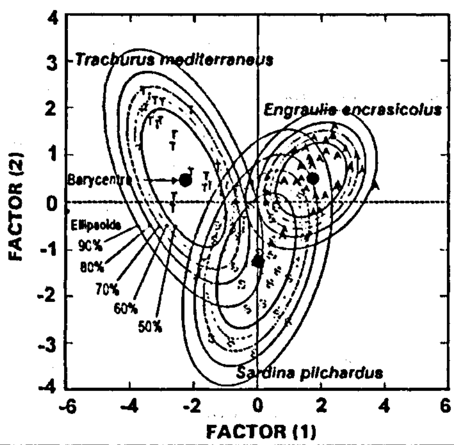
Figure 3. *Radar plot of the mean percentage of change of the 3 output values of a Sardine school by varying all the input values +10% - 10%. (Experiment with 20 parameters)*

An other way to test the network, but especially to study the impact of each variable on the output, is the use of a certain option of SCHOOLBRAIN to vary all input values by a small amount, and see the result on a certain output (Figure 3). Through such tests, we can also export useful information about the sensitivity of the trained network.

3.3. Discussion

The same data that we used for SCHOOLBRAIN, were submitted in Discriminant Function Analysis

Figure 4. *Scatterplots of the 3 species in the plan of the two discriminant functions. (Factor 1 and Factor 2)*



	Anchovy	Sardine	Horse m.	TOTAL
ANCHOVY	97	28	2	127
SARDINE	65	35	1	101
HORSE-M	7	17	18	42
TOTAL	169	80	21	270

Table 2. *Two-way table to demonstrate accuracy of classification of the multiple linear regression model of the DFA. Groups (rows) by Predict (columns)*

(DFA). The results of a multiple linear regression model were very poor (Table 2), because the data did not satisfy the multivariate normality in the distribution of the parameters. Thereby we can see significant overlaps in their multiple regression scatterplots, which eliminate the discrimination (Figure 4). Instead of these problems, generally DFA has the advantage over ANN, of assigning a probability level to each case for its group membership. Our experiments suggest that we can introduce such a concept in an ANN, if we assign the value of 1 to the neuron of the group of the training case, and the value of zero to the other neurons.

4. REFERENCES

1. Diner, N. et al. INES-MOVIES: a new acoustic data acquisition and processing system. ICES CM 1989/B:45, Fish Capture Committee.
2. Gerlotto, F. and Freon, P. School structure observed in some clupeid species. ICES/FAST Work Groups, Ostend, Belgium, 19-22 April 1988.
3. Georgakarakos, S. and Paterakis, G. "SCHOOL": A software for fish-school identification. ICES 1993/B:8 Fish Capture Committee.
4. Lawrence, J. 1993. Introduction to Neural networks. Design, theory and applications. California Scientific Software Press, Nevada City, pp 324.1993.
5. Scalabrin, C., Weill, A., Diner, N. The structure of multidimensional data from acoustic detection of fish schools. European Conference on Underwater Acoustics. 1992 ed. by M. Weydert, Commission of the European Communities, Brussels.