ICES Journal of Marine Science, 53: 173–180. 1996

II. Classification and identification

Artificial neural networks as a tool for species identification of fish schools

J. Haralabous and S. Georgakarakos

Haralabous, J. and Georgakarakos, S. 1996. Artificial neural networks as a tool for species identification of fish schools. – ICES Journal of Marine Science, 53: 173–180.

Fish schools of sardine, anchovy, and horse mackerel can be discriminated from each other, under given conditions, using a set of parameters extracted from echo-integration data. Trawl sampling and hydroacoustic data were collected in 1992 and 1993 in the Thermaikos Gulf by using a towed dual-beam 120 kHz transducer. The parameters extracted from the available schools were used to train multi-layered feed-forward artificial neural networks. Various applied networks easily generated associations between school descriptors and species identity, providing a powerful tool for classification. The expertise of the trained network was tested with data from identified schools not used in training. The use of neural networks cannot replace classical statistical procedures, but offers an alternative when there are significant overlaps in the school characteristics and the parametric assumptions are not satisfied.

Key words: backpropagation, classification, echogram, feed-forward, fish school, hydroacoustics, identification, neural networks, pelagic fish, supervised learning.

J. Haralabous, and S. Georgakarakos: Institute of Marine Biology of Crete (IMBC) – Hydroacoustics, PO Box 2214 GR-71003 Heraklion, Crete, Greece. Correspondence to Haralabous [tel: 030 81 243368, fax: 030 81 241882].

Introduction

The development of fish-species identification techniques based on hydroacoustic data is one of the keys for the reduction of error in biomass estimation. Echogram scrutinizing methods, combined with concurrent trawling data and human experience 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 sufficiently describe the structure of particular fish aggregations (Diner *et al.*, 1989; Georgakarakos and Paterakis, 1993) or acoustic populations (Gerlotto and Fréon, 1988) and support the species identification procedures.

The selection of the best descriptors, and the accuracy of classification, are the two main associated problems. Classical statistical methods, such as PCA (principal components analysis) and DFA (discriminant function analysis), are the most commonly performed techniques in this area (Scalabrin *et al.*, 1992). However, the strict constraints limit the reliable use of the above procedures in some cases.

Feed-forward artificial neural networks (ANN) do not demand any assumptions concerning the distributions. They are now widely used for regression, classification, and discrimination, though their use is rather new in fish school identification and classification problems. The aim of the present study is to discuss features of neural networks which could serve as a tool for classification, and to present neural network applications for the species identification of small pelagic fish schools.

Materials and methods

Data acquisition and extraction of descriptors

School data were obtained from hydroacoustic surveys with the RV ''Philia'' in the Thermaikos Gulf in 1992 and 1993. Data were collected using BioSonics dual-beam equipment operated at 120 kHz. The pulse duration was 0.5 ms and integration was carried out over 1 m intervals through the water column. The data have been analysed by using ''School'' software, developed in IMBC (Georgakarakos and Paterakis, 1993), for school identification and extraction of the required parameters. The software includes the following routines.

''*SCROLL*'': The formation of the echogram is based on a set of elements (pixels) with a resolution equal to 1 m on the vertical axis and to 1 ping interval on the

1054–3139/96/020173+08 $18.00/0

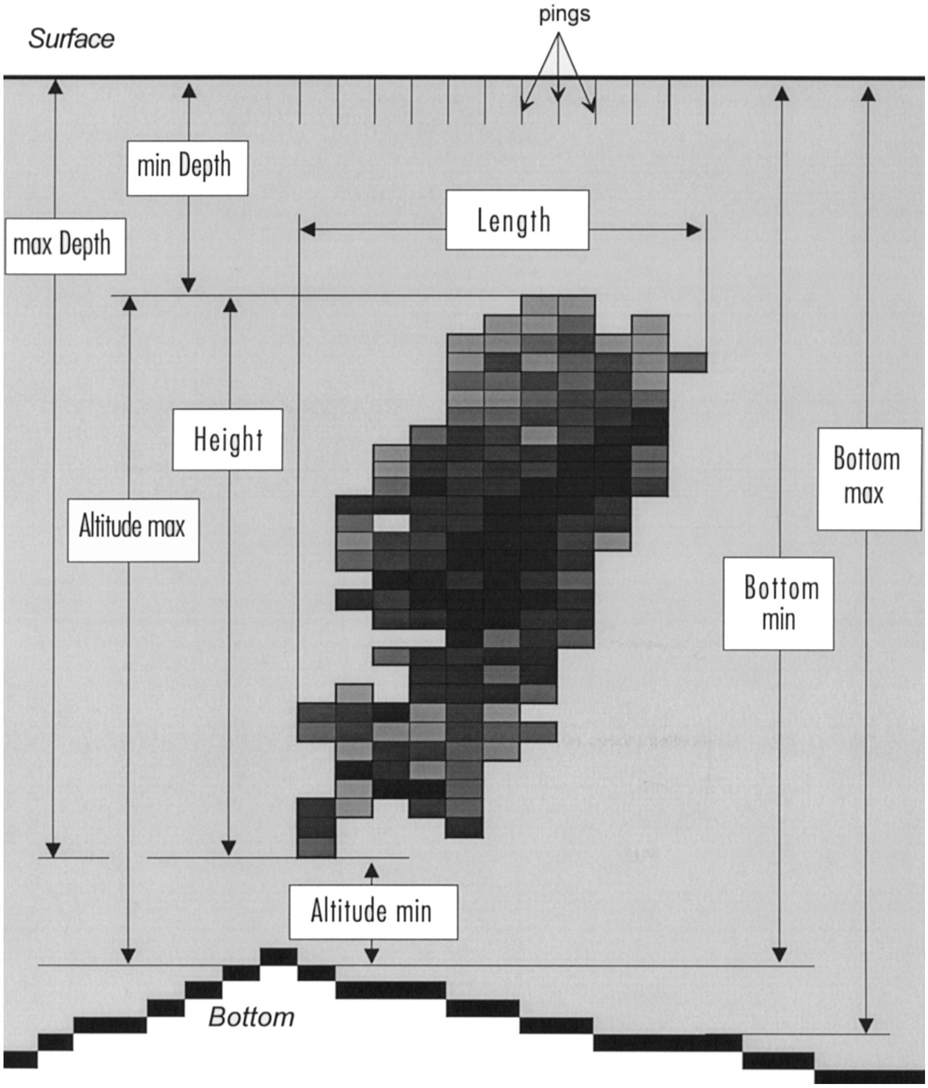

Figure 1. Diagram of a digitized school echogram with indications of some morphological and bathymetric descriptors. Grey scaling of pixels corresponds to energy scale.

horizontal axis (Fig. 1). The area of each element is defined as the product of the horizontal distance between two successive pings in metres and the vertical distance between two successive integration layers.

''*FILTER*'': Three procedures have been used to perform school recognition: (i) an echo-integration threshold in order to cut off very low biomass concentration, (ii) an algorithm that detects contiguous elements along the same ping and/or contiguous elements from one ping to the next. Elements that fulfil this continuity test are considered as belonging to the same aggregation, (iii) a school mean energy threshold to remove plankton aggregations from the analysis of schools. Horizontal dimensions were corrected by taking into account the school depth and the nominal beam angle of the transducer (MacLennan and Simmonds, 1992). Schools with estimated negative length were rejected.

''*SCHOOLBASE*'': Database used for training purposes. Up to now, 3420 schools have been encountered, digitized and analysed with this software. More than 90% of these schools belong to the three most common species in this area: sardine (*Sardina pilchardus*), anchovy (*Engraulis encrasicolus*), and horse mackerel (*Trachurus trachurus*). For 762 of the above schools (22.3%) we acquired information by trawling. We have chosen those schools detected during trawling and when the catch was monospecific. However, only 270 of these 762 schools (35%) were identified with the highest degree of certainty. From these 270 schools we

Table 1. The main school descriptors used in the neural network application grouped by category. The asterisk indicates descriptors used in the form of summary statistics (lowercase letters indicate the statistic, for example, min = minimum, max = maximum, var = variance, cv = coef. of variation).

School descriptor	Symbols and computations	Units
General		
Species ID	SPE	
Morphological		
Height	H	m
Length	L	m
Perimeter	P	m
Area	A	m^2
Elongation	ELON = L/H	
Circularity	CIRC = $P^2/4\pi A$	
Rectangularity	RECT = (LH)/A	
*Radius of perimeter	Rmean, Rmin, Rmax, Rcv	m
Fractal dimension	F = 2[ln(P/4)]/ln(A)	
Bathymetric		
*School depth	Dmean, Dmin, Dmax	m
*Bottom depth	Bmean, Bmin, Bmax	m
*Altitude	Amean, Amin, Amax	m
Energetic		
Total school energy	E	V^2
*School energy	Emean, Emax, Ecv	V^2
Index of dispersion	I = Evar/Emean	V^2

used a subset of 140 that were sampled under similar environmental conditions (same surveyed area, time of day, season of year, distance from land, etc.) to avoid variability in data, due to factors that have not been used in the analysis. More than 25 parameters could be calculated by ''School'' software classified into three main groups: morphological, energetic, and bathymetric (Table 1).

Neural network application

An artificial neural network (ANN) uses a highly interconnected group of simulated neurons (units, elements, or nodes) that process information in parallel. Its main concept is to approximate functions with raw sample data, i.e. to approximate input–output responses learning from experience. A review of the current development of ANNs and their applications can be found in Rumelhart *et al.* (1986), Kosko (1992), Hecht-Nielsen (1991), Lawrence (1993), and Ripley (1993).

From the broad range of ANNs, we are involved here with pattern classifiers that use supervised learning algorithms. The patterns (in our case the values of the school descriptors) are presented to the network, which is supposed to categorize them according to the predefined classes (in our case the three species IDs: anchovy, sardine, or horse mackerel). A supervisor presents several patterns together with their correct classification. After having been trained, the network should be able to

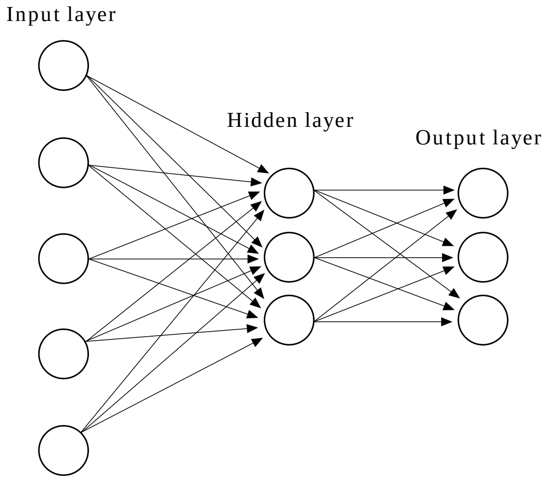

Figure 2. A generic feed-forward neural network with a single hidden layer.

correctly classify patterns different from those used during the training phase. The output of the network can be seen as a mapping from the input vector space to the output vector space (Ripley, 1993). Learning is taken here as the means by which such mapping can be constructed.

In the present study we used the most common structures for supervised learning: the multi-layered feed-forward networks. Neurons are organized in three types of successive layers: the input layer that here retains the school echogram descriptors, the hidden layer(s), and the output layer that holds the classification results, while neurons in a given layer do not connect to each other. They take their inputs only from the previous layer and send their outputs only to the next one, thus computing a result very quickly (Fig. 2). The neurons are connected through a set of connection weights, or synaptic weights that multiply the corresponding input signal. During training, the network takes every school as separate input (one training fact at a time) and produces an actual output pattern (species' ID) according to a sigmoid transfer function. Before taking the next fact, it compares this output with the desired one, calculating the mean squared error (MSE). According to Ripley (1993), the weight parameter is chosen to minimize the MSE, as would be done in a non-linear regression. This is a general minimization problem and the most widely used algorithm applied is the generalized delta rule (GDR); the networks that use this rule are called back-propagation networks (Hecht-Nielsen, 1991). The delta rule states that if there is a difference between the actual output pattern and the target output during training, the weights are changed to reduce the difference. The term ''back-propagation'' comes from the fact that correction signals propagate back through the network during training.

Table 2. Classification results of discriminant function analysis (DFA).

Actual group	No. of cases	Predicted group membership		
		Anchovy	Sardine	Horse mackerel
1. Anchovy	75	71 (94.7%)	4 (5.3%)	0 (0.0%)
2. Sardine	40	10 (25.0%)	30 (75.0%)	0 (0.0%)
3. Horse mackerel	25	0 (0.0%)	1 (4.0%)	24 (96.0%)
Total	140	81	35	24

Percentage of ''grouped'' cases correctly classified 89.29%.

At IMBC we have used two different commercial neural network simulators to develop the specific ANNs applications that can classify the species of small pelagic fish from their school echograms. The first one (ANN1) is based on the ''BrainMaker Professional'' (California Scientific Software©) simulator, while the second (ANN2) is based on the ''NeuroShell II'' and ''Neuro Windows'' Dynamic Link Libraries (Ward Systems Group©). The training is based on subsets of data for which the species identity is known by means of trawling. Different subsets of the known data are used for testing. Their detailed topology is described and discussed in the Results. The dynamic of the above ANNs can be modified by the researcher to meet different sets of learning parameters, e.g. training tolerance, learning rate, noise thresholds, ability to add hidden units if necessary during training, and smoothing factors. Especially in ANN2 the classic back-propagation algorithm has been improved by the momentum correction term (Rumelhart *et al.*, 1986) which is an exponential smoothing, appropriate when the examples are presented in a random or unstructured order in the network.

Results

Preliminary data analysis

Data were first analysed with classical statistical procedures using commercial statistical software (SPSS for Windows Release 6.0). Discriminant function analysis (DFA) was performed using multiple linear regression models. A stepwise variable selection method was chosen and two discriminant functions were computed for discrimination among the three species. Table 2 contains the classification results of DFA, and Figure 3 the discriminant plot.

The essential problem encountered with the application of DFA (Haralabous and Georgakarakos, 1993) was that the main assumptions of multivariate normality of the data set and the equality of their covariance matrices were not satisfied. Box's M-test of equality of group covariance matrices showed that the above hypothesis must be rejected ($p<0.001$). Significant overlaps in school features (see the discriminant plot) indicated the low discriminating power of the functions. This is also obvious in the relatively poor classification results in Table 2.

Neural network topology and training

The input and output layer structure of the ANN1 and ANN2 is relatively simple; there are as many units (neurons) in the input layer as the number of input descriptors, and as many units in the output layer as the number of different species. Each output unit, corresponding to a certain species, takes the value of 1 if the school echogram belongs to this species, otherwise the value of 0.

The structure of the hidden layer(s) varies depending on the equilibrium of some other factors, such as number of examples, duration of training, accuracy of classification, training tolerance, etc. A general rule of thumb is to start with a number of hidden units slightly fewer than half the number of inputs and outputs (Lawrence, 1993) and to continue adding hidden units depending on the classification results of both training and testing sets. Preliminary practice suggests that as the number of hidden neurons approaches the number of patterns in the training set, the danger of memorization increases, i.e. the network may end up memorizing the facts rather than learning to generalize about them; thus, the network may train very well but test poorly. Here the experiments suggest keeping the number of hidden units in the range 12 to 18.

For a certain total number of hidden units, the decision of choosing the number of hidden layers and the number of neurons in each of those hidden layers was based on the Fiesler's maximum interconnection topology theory (Fiesler, 1993). In ANN1 there is a single hidden layer in the beginning, while during training the system can dynamically add new hidden layer(s) according to the process of successful classifications. In ANN2 we used two different network topologies: one with a single hidden layer (ANN2-a) and a second with two hidden layers (ANN2-b) for comparison. Table 3 summarizes the topology used for ANN1 and ANN2.

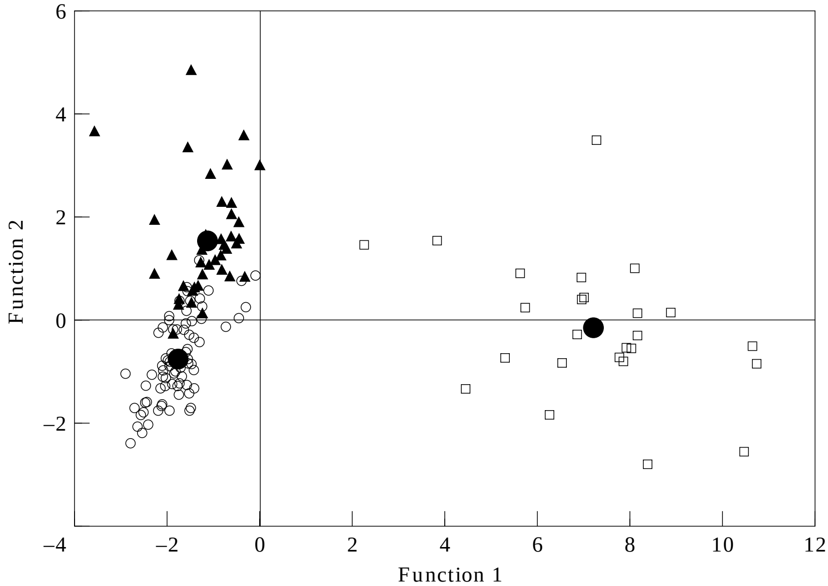

Figure 3. Plot of the three species data on the two discriminant functions obtained from DFA. ● = Group centroids; □ = Horse mackerel; ▲ = sardine; ○ = anchovy.

Table 3. Layer structure of neural networks ANN1 and ANN2. The last column contains the correct classification results of test sets in the range 5% to 35% of the total available data.

Topology	Network	Number of neurons per layer				Correct test results
		Input	Hidden1	Hidden2	Output	
1	ANN1-a	25	14	—	3	89–96%
2	ANN1-b	25	12	3	3	88–97%
3	ANN2-a	25	18	—	3	94–100%
4	ANN2-b	25	13	2	3	92–98%

In Figure 4 we can see the progress of the ANN1 training as a function of the errors of classifications on the number of training cases in each output unit, separately and overall. After some training rounds (each of which involves a complete pass of 119 used cases) the network could easily correctly classify school echograms of horse mackerel. The correct classification of sardine and anchovy school echograms needed almost five times more training rounds. The same reaction has been noticed using ANN2.

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 was positively connected with the accuracy of predictions. Generally, the number of training rounds was larger when we used two hidden layers (ANN2-b) because of the larger number of weights needed.

Testing and accuracy of classification

To measure the classification accuracy of the networks, we need to compare the actual output of the network with the correct output over a number of testing trials. This requires additional randomly selected examples beyond those used for training the networks. The most widely used method for obtaining this test set is to reserve a separate representative subset of the available examples. These example patterns must be new because, if we use the same examples for training as are used for testing, all we are determining is how well the networks learned the training patterns. What we are really interested in is how well the networks had learned to approximate the ‘‘mapping’’ function for arbitrary input values (Hecht-Nielsen, 1991).

In the present study we experimented with testing subsets ranging from 5–35% of the available data. Depending on other learning settings, the testing

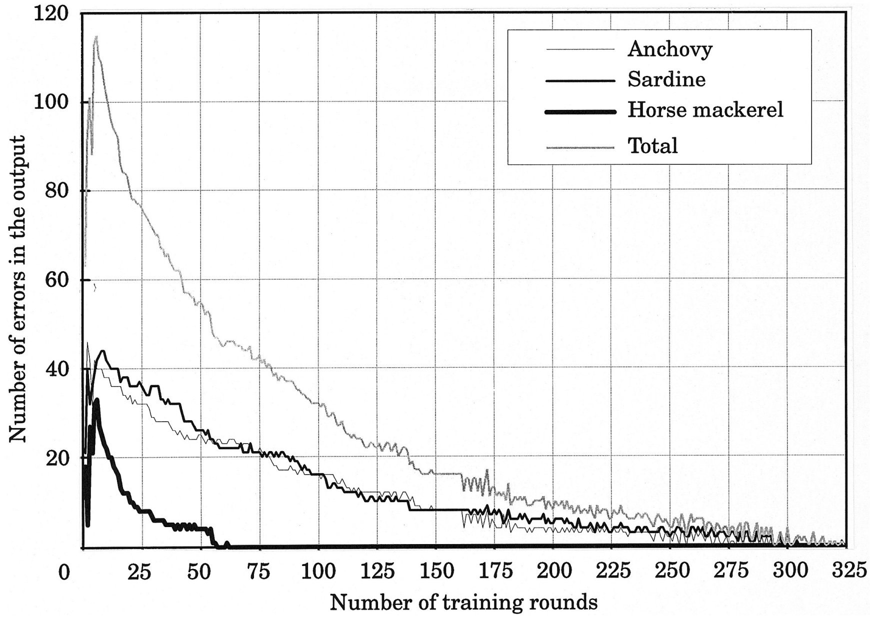

Figure 4. The progress of the neural network application ANN1.

Table 4. Classification results of a test set of 49 examples (35% of the available data set) using the topology 3 (ANN2-a).

Actual group	No. of cases	Predicted group membership		
		Anchovy	Sardine	Horse mackerel
1. Anchovy	23	22 (95.65%)	1 (4.35%)	0 (0.00%)
2. Sardine	18	1 (5.55%)	17 (94.44%)	0 (0.00%)
3. Horse mackerel	8	0 (0.00%)	0 (0.00%)	8 (100%)
Total	49	23	18	8

Percentage of ‘‘grouped’’ cases correctly classified 95.92%.

predictions were correct for 88–100% of the testing cases. The best testing performance was given by the ANN2-a and it was in the range of 94–100% (Table 3). In Table 4 we can see the detailed classification performance of the above network for each species if we use a test set of 35% of the available data.

Contribution factors

The contribution factor of a variable is the sum of the absolute values of the weights leading from this particular variable. In Figure 5 we can see the Pareto chart with the contribution factors of the first 15 descriptors in ascending order given by the ANN2-a. The order was the same for these 15 descriptors in all topologies but, after the 15th, the order was different in the various configurations. Despite the absolute value, which differs slightly even for the first 15 common-ordered variables, their order indicates that the heaviest impact in classification is assigned to the morphological and bathymetric features, while the energetic properties of a school come next.

Discussion

The problem of species identification of fish schools using hydroacoustic data is approached in the present study as a problem of mapping school echogram features on species IDs. The morphological, bathymetric, and energetic characteristics used here do not restrict the impact of other recorded features, which may possibly be better for the classification mapping. Comparative experiments are needed for the evaluation of the best sets of features. For the present study, a comparison is made on a set of school echogram descriptors of three species within the range of the explanatory limits

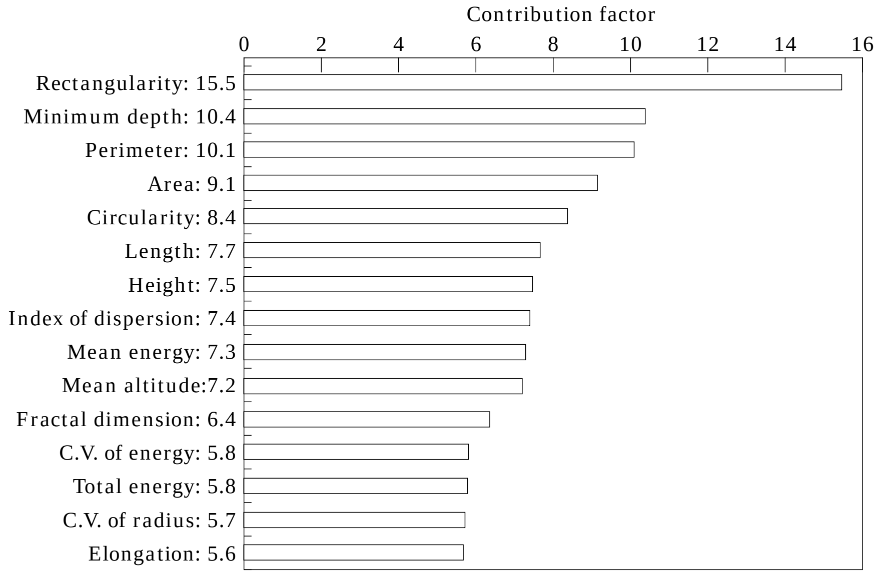

Figure 5. Pareto chart of the contribution factors of the first 15 descriptors (ANN1-a).

(concerning the ecological characteristics of the certain sampling area, the spatio-temporal features of sampling, the monospecific composition of schools, the instrument and software used, etc.). The specific results on the impact of each descriptor variable must also be seen under the above limitations, as well as the size and species composition of the sample.

The small amount of available data affects the validation of the network used and its accuracy. As is often noticed, there are probably enough data to train the network, but not enough to hold out for validation and acceptance test sets. One of the ways to decide how large the test set should be is to try progressively larger test sets until the MSE starts to converge to a fixed value. This value is one of the most important measures of error, and is thus essential for the validation of the network. Starting from 5% of the available examples in our experiments, we noticed that the MSE did not clearly converge to a fixed value even when about 55% of the total examples had been used, obviously because of the progressive removal of necessary cases from a small training set. In these instances, one option that can help is the ''leaving-one-out'' method proposed by Hecht-Nielsen (1991): if there are N examples available, then train the network N different times using $N-1$ of the examples, each time excluding a single different example as a singleton test set. On each of these N trials, the training set is used for testing the network during training, and training is ended when the error curve levels out. Then the error of the network on the held-out test example is measured. After doing this N times the mean of the squared errors made on the held-out examples is calculated. This is one of the ways to estimate the overall network performance that would be achieved if more data from the real environment were available. In our study this value was found to be at 0.071 for the ANN2-a configuration. This means that there is an expected error of 7.1% for each output neuron (recall that the output values have a range between 0 and 1).

It must be noted that a comparison between neural networks and DFA is not straightforward because: (a) an ANN can be tested only on a subset of training-free cases, while DFA can acceptably be tested on the whole data set. A proper comparison should take a single subset of the data to train the ANN and DFA, and then a separate subset to test both the ANN and DFA; (b) DFA recommends the absence of correlation among descriptors, suggesting techniques to ensure this directly (for example through selection of descriptors ranked by correlation), or indirectly (in combination with PCA and the use of the principal components as descriptors). In the applied ANNs, however, we could use all the descriptors without being worried about correlations, leaving the ANN to support an impact factor for each of them; (c) the high number of descriptor variables for the above three species implies that a much larger sample than the available 140 cases is needed for a proper DFA. This is a fixed limit for the number of descriptors that can be tested at once in a DFA, while in the ANNs it affects only the convergence of the mean squared error.

The main reason for artificial neural networks being potentially useful for classification problems is their property of being classifiers which construct non-linear functions from inputs to targets. Most classical statistical techniques that could be used instead of ANNs are severely constrained if linear regression models have to be used. During the 1980s, the proliferation of computer power removed the constraints of linearity and a number of non-linear approaches have been developed for regression and discrimination, often referred to as semi-parametric methods. According to Ripley (1993) ''both feed-forward neural networks and semi-parametric statistics are 'black-box' approaches; they provide a prediction for any input, but no readily interpretable explanation for that prediction. As such they lose some of the power of linear statistical models in statistical inference''. However, many approaches are now being combined to explain the basis for the predictions made by neural networks.

Acknowledgements

This study was partially financed by the Commission of the European Union within the framework of the project ''Biomass''. The authors thank their partners for the excellent collaboration, especially E. J. Simmonds for encouragment throughout the ANN implementation.

References

Diner, N., Weill, A., Coail, J. Y., and Coudevill, J. M. 1989. ''Ines-movies'': a new acoustic data acquisition and processing system. ICES CM 1989/B: 45.

Fiesler, E. 1993. Connectivity maximisation of layered neural networks for supervised learning. *In* Proceedings of the International Conference on Artificial Neural Networks. Ed. by S. Gielen and B. Kappen. Springer-Verlag, London. 514 pp.

Gerlotto, F. and Fréon, P. 1988. School structure observed in some clupeid species. ICES/FAST Work Group, Ostend, Belgium, 19–22 April 1988.

Georgakarakos, S. and Paterakis, G. 1993. ''School'': A software for fish-school identification. ICES CM 1993/B: 8.

Haralabous, J. and Georgakarakos, S. 1993. Fish-school species identification using a neural network. ICES CM 1993/B: 9.

Hecht-Nielsen, R. 1991. Neurocomputing. Addison-Wesley Publishing Company, USA. 433 pp.

Kosko, B. 1992. Neural networks for signal processing. Prentice-Hall, New Jersey. 399 pp.

Lawrence, J. 1993. Introduction to neural networks. Design, theory and applications. California Scientific Software, Nevada City. 324 pp.

MacLennan, D. N. and Simmonds, E. J. 1992. Fisheries acoustics. Chapman and Hall, London. 325 pp.

Ripley, B. D. 1993. Statistical aspects of neural networks. *In* Chaos and networks – statistical and probabilistic aspects, pp. 40–123. Ed. by O. E. Barndoff-Nielsen, D. R. Cox, J. L. Jensen, and W. S. Kendall. Chapman and Hall, London.

Rumelhart, D. E., McClelland, J. L., and the PDP Research Group 1986. Parallel distributed processing: explorations in the microstructure of cognition. MIT Press, Cambridge, Massachusetts.

Scalabrin, C., Weill, A., and Diner, N. 1992. The structure of multidimensional data from acoustic detection of fish schools. *In* European Conference on Underwater Acoustics, pp. 141–146. Ed. by M. Weydert. Elsevier Applied Science, London.