

Biological pattern recognition by neural networks

R. Simpson¹, R. Williams², R. Ellis¹, P. F. Culverhouse³

¹Department of Psychology, Polytechnic South West, Plymouth PL4 8AA, Devon, UK

²Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth PL1 3DB, Devon, UK

³School of Electronic Communication and Electric Engineering, Polytechnic South West, Plymouth PL4 8AA, Devon, UK

ABSTRACT: Neural network analysis was proposed and evaluated as a method for image analysis of plankton data derived from automatic counting techniques. It was shown that a neural network with 2 layers of weights was capable of learning a large data set by the backward-error propagation method. Significant results were achieved in separating novel images of 2 co-occurring species of *Ceratium* spp. from the western North Atlantic Ocean.

INTRODUCTION

There is a growing concern amongst the marine scientific community about the diminishing numbers of competent plankton taxonomists remaining in, and being recruited to, research. This has led to a number of initiatives attempting to overcome this problem using automated pattern recognition techniques (Jeffries et al. 1984, Rolke & Lenz 1984, Steidinger et al. 1990). Certainly it is a generally held belief that the majority of plankton samples will eventually be sorted, counted and taxonomically analysed using automatic techniques. No definitive system has yet been built but it is highly likely that pattern recognition and image processing will be the method to achieve the processing of the data. There is also a growing requirement for discrimination and identification of video images collected from cameras mounted on remote vehicles and towed systems. Recognition of biological patterns by neural networks is a new concept and one which we suggest could play a role in achieving these ends.

The majority of previous attempts to use automated pattern recognition for taxonomic identification have utilised explicitly statistical techniques, whereby morphological measurements are made on sample data (either by hand or by some automatic image analysis technique), and the measurements submitted to discriminant analysis (Yarranton 1967, Jeffries et al. 1980, 1984). Typically such techniques involve making a wide range of measurements, and using the statistical analysis to determine which measurements or combi-

nation of measurements allows the best discrimination to be made. Obtaining images clean enough to enable accurate measurement is itself quite problematic. A neural network approach was considered suitable for this type of problem for a number of reasons: (1) neural networks (strictly, artificial neural networks) are able to be taught, i.e. exposed to data and 'told' what the correct response (e.g. category) is for that stimulus, and do not need to be given a rule-base to determine behaviour; (2) they are able to form arbitrary input-output mappings, and so can be used in a wide variety of domains; (3) they have been shown to be largely resistant to noisy or imperfect inputs; (4) their massive operational parallelism means that hardware versions of networks can be made that perform very complex mapping tasks at great speed.

MATERIALS AND METHODS

Samples. To assess the possibilities of artificial neural networks' discriminating images of planktonic species, specimens of the genera *Ceratium* were used. *Ceratium* spp. are generally compressed in one plane and are ideal organisms for drawing or video imaging. They are also morphologically variable and specimens are at times difficult to separate, when there can be up to 20 different and distinct species of the genera occurring in one plankton sample. The 2 species selected were *C. arcticum* (Ehrenberg) Cleve, and *C. longipes* (Bailey) Gran., which have sympatric geographical distributions in the western Atlantic Ocean.

Fig. 1a, b shows 4 abundance levels (where the sampled area is bounded by a solid line on the maps). The analysts from the Continuous Plankton Recorder (CPR) team (Glover 1967, Colebrook 1982) identify phytoplankton from samples collected in the western Atlantic, and thus are making taxonomic decisions differentiating these species. Samples of phytoplankton were used from this ongoing survey, and assistance in classifying these samples was obtained from the CPR personnel and used to test the efficacy of this neural network technique.

Neural networks. Neural networks consist of large numbers of processors ('nodes'), massively interconnected by variable connections ('weights') (see Fig. 2). Typically research in this domain is done on serial digital computers simulating the parallel operation of these networks, as is the case here.

The neural network model used in the experiments reported here is a *feed-forward activation* model (meaning that the nodes are arranged in layers, and activation of the nodes feeds forward from the input layer to the output layer, possibly through intervening lay-

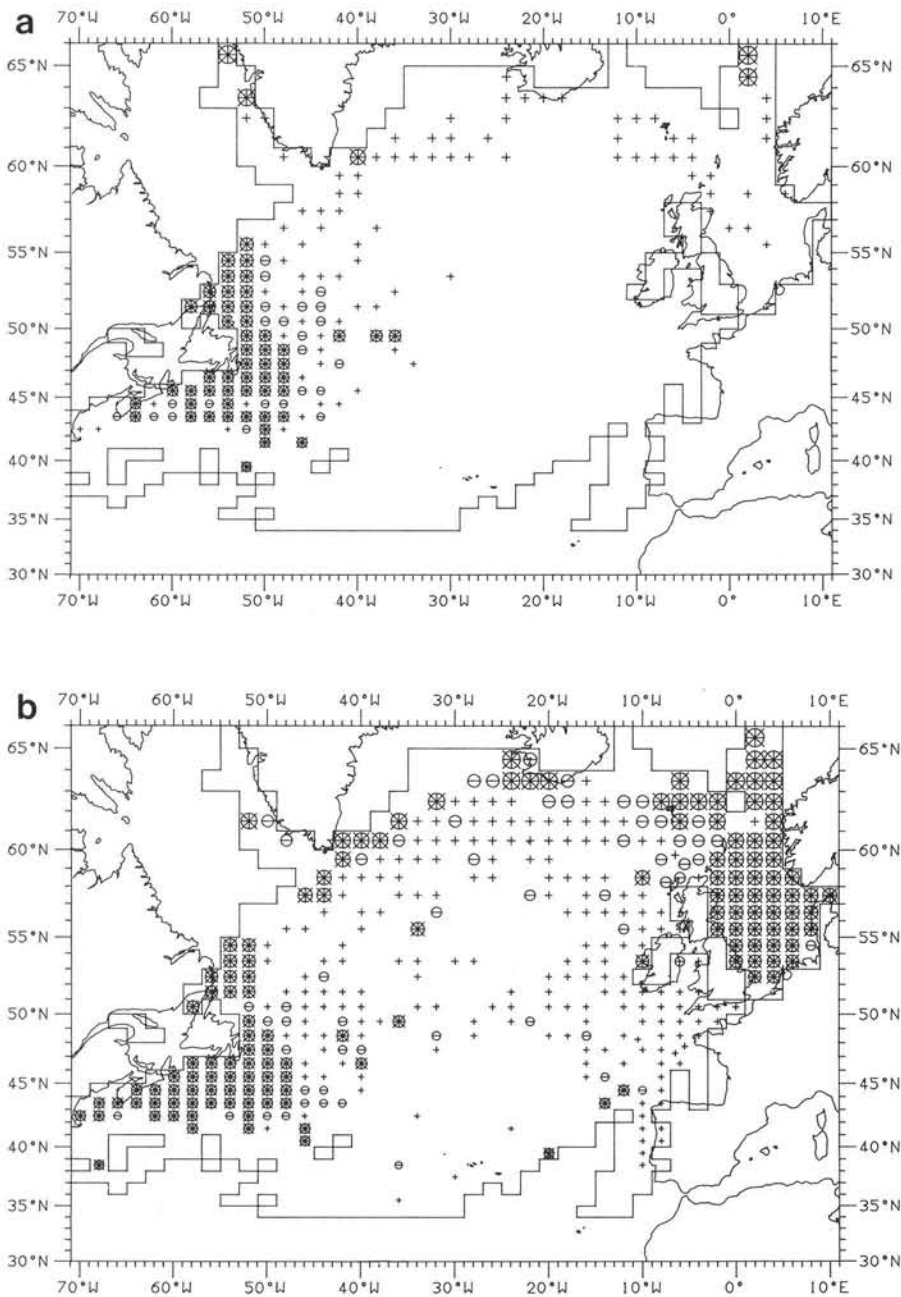


Fig. 1. *Ceratium* spp. Annual mean distribution and abundance of (a) *C. arcticum* and (b) *C. longipes* in North Atlantic, from sampling at 10 m depth by the continuous plankton recorder. Data for all months sampled were combined to show mean abundance from 1958 to 1989. Samples were assigned to rectangles 1° lat. by 2° long. Rectangle means are represented by 3 graded symbols (abundance: +, least; ⊕, intermediate; ⊗, most). The absence of symbols (in the sampled area) indicates that the species were not found in the CPR samples. The boundary of the sampled area is shown by straight lines

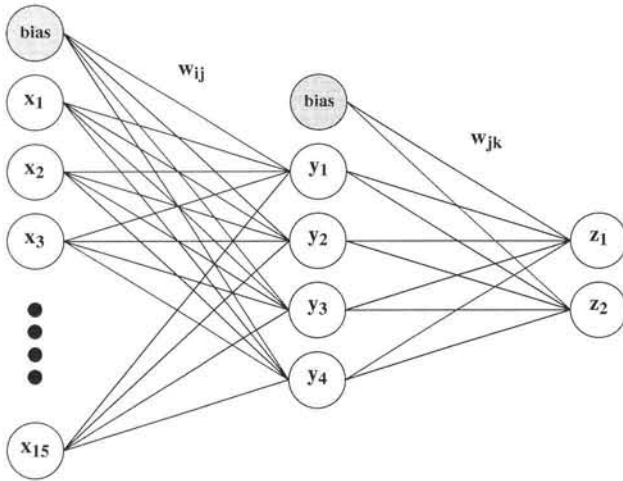


Fig. 2. A feed-forward network with 15 input units, 4 hidden units and 2 output units. w_{ij} indicates a weighted connection from node i to node j

ers), taught by the *backward error propagation* technique (Rumelhart et al. 1986). A network is exposed to an input pattern, and the weights are adjusted so that the required output pattern results at the output layer. In addition to the nodes described, there is a *bias* node in each layer prior to the output layer, providing a stimulus to each layer forward of the input layer. This is required in order to ensure the dynamic range of the input to a node is within its operational bounds. The backward-error propagation (or 'back-prop') algorithm

does this iteratively, repeatedly exposing the network to the input-output pattern pairs until the amount of operational error shown by the system reaches a sufficiently low value. This being so, when shown a pattern in the training set, the network will respond appropriately. The strengths of the weights connecting the input and output nodes can be said to represent the categories that the network has learnt. If there are additional layers of nodes (called hidden units in that they are not directly accessed by inputs or outputs) they can be said to have formed an internal representation. The real strength in this approach is not in teaching a network to categorise a particular set of data, but in using a network so taught to categorise new, novel data. If a network is able to respond appropriately to novel data (or test data) it can be said to have *generalised* from the training set. This is in effect how humans operate. They are shown repeated examples of input-output pairs, (somehow) come to learn them, and extend their knowledge of these examples to new cases. Experiments were done in order to investigate the possible generalisation ability of networks in this domain.

Obtaining the images. Drawings of 110 specimens of the dinoflagellate genus *Ceratium* (*C. arcticum* and *C. longipes*), removed from CPR samples, were obtained. These were made either from line drawings taken with a camera lucida fitment on a microscope or directly from the microscope slide using a video camera and digital frame grabber (Fig. 3a). The line drawings were then digitised, and thresholded to give a binary black

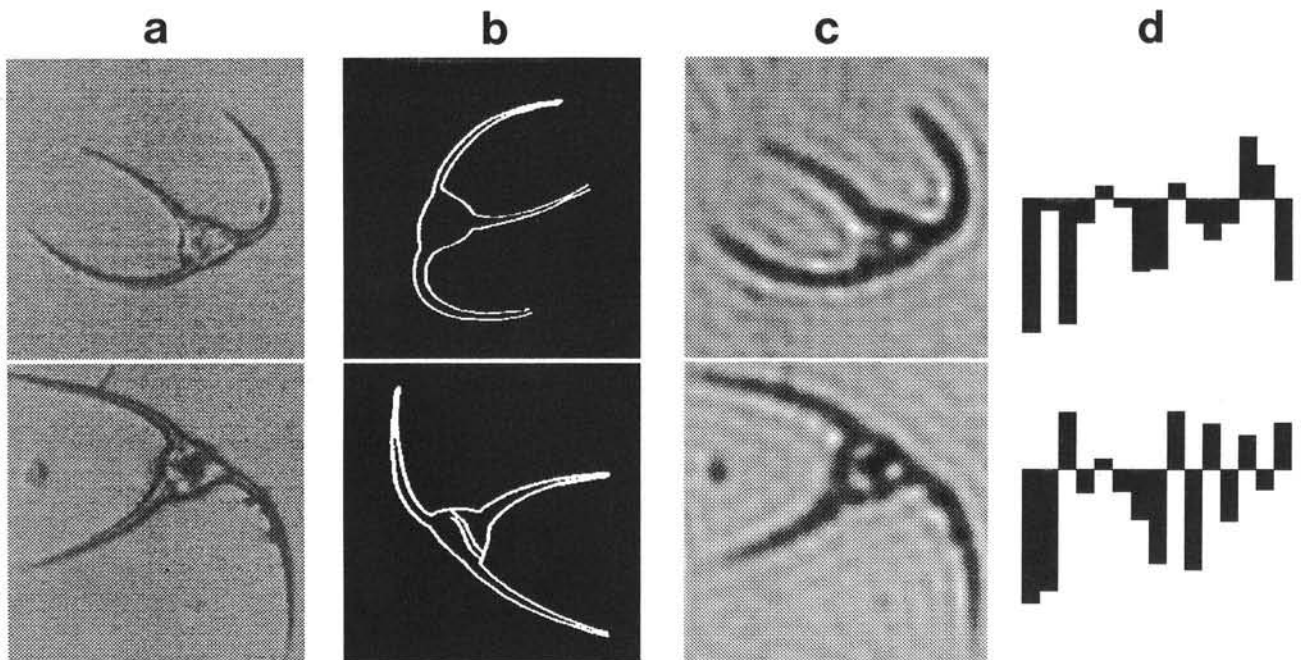


Fig. 3. (a) Grey-level images of *Ceratium longipes* (upper) and *C. arcticum* (lower). (b) Binary line drawings of same. (c) Reconstituted image of only the lowest 16 spatial frequencies from (a). (d) Frequency gradient histograms of the images in (c)

and white image (Fig. 3b). These data were initially thought to consist of roughly half *C. longipes* (Bail.) and half *C. arcticum* (Ehrb.). Given the generally difficult nature of *Ceratium* classification it was thought inadvisable to obtain a cast-iron set of classifications from one expert, so a technique was devised for obtaining the 'correct' class of each specimen. A randomised double set of images was produced, and a number of plankton analysts with a wide range of classificatory experience in this field were asked to identify each image. Each set of images given to each individual were randomised differently to prevent the occurrence of order effects.

The results of the analysts' identifications were analysed in 2 ways. Firstly, each individual's classifications were analysed to see how self-consistent they were, that is, how reliably a particular individual classified images to be a particular species. This could be calculated since each image occurred twice in the set of images being classified. This was the measure of *internal consistency*. Secondly, the results were analysed to determine the *external consistency* of the group of individuals. Table 1 shows the percentage agreement between each individual and every other individual.

By ordering the results by internal consistency it can be seen that the 3 most internally consistent set of classifications (A, B and C) were mutually consistent to a high degree. Beyond this group both internal and external consistency dropped off sharply. Working on the premise that if one is largely right, one's decisions will be largely consistent with someone who is also largely right, whereas if one is wrong one may be wrong in very different ways to another person who is wrong, we suggest that this clustering of the more consistent performers indicates that they are all largely right (i.e. right thinkers must think alike, wrong thinkers needn't). Given this we can use their judgments to give us a 'definitive' benchmark classification

for each species. Where there are disagreements within this group a simple majority decision was taken as the class. In a few cases the majority decision was considered to be insufficiently overwhelming, and that image was discarded for these experiments. Eventually a set of 100 images, 50 *Ceratium longipes* and 50 *C. arcticum*, was arrived at, each one classified as such more than 80 % of the time by these 3 experts. This set constituted our image data.

In order to reduce the size of the input to the network some pre-processing of the image was done. In all experiments the frequency power spectrum of the image was obtained by performing a Fourier Transform (FT) and summing all the information for each frequency. This gives us a set of frequency responses for the image. The FT was chosen as a suitable pre-processing technique for several reasons. Firstly, the nature of the FT means that the plankton in the image need not be positioned or centred prior to the transform, since the result of the transform is largely position invariant with respect to large detail in the image. This saves us quite a problem, since techniques requiring centering of an object can be very sensitive to slight errors in positioning. Secondly, an FT can be performed in a parallel processing fashion on suitable hardware, thus enabling a fast transform, and possibly the development of a single network incorporating the pre-processing and the later processing. Thirdly, a power spectrum can be obtained from an FT, and a frequency histogram from the power spectrum. There is a very great compression of information from the grey-level image to the frequency histogram, and this compression has advantages for a network simulation, namely of network size, giving faster simulation times, and of learning speed. Simpson (1989) has previously reported on experiments that show that low spatial frequencies contain sufficient information to discriminate between images of angles of different sizes. Since it appears that angular data play a role in identifying *Ceratium* species, spatial frequency data can be regarded as being both efficient and meaningful.

It had been observed that the gross structure present in the lowest 16 spatial frequencies was sufficient for a *Ceratium* specimen to be identified by an expert. Reconstructing an image from an FT from which higher frequencies have been removed results in a somewhat blurred and wavy image (Fig. 3c). In the case of *Ceratium* it was observed that such a reconstructed image could be classified by an expert with ease. The higher frequencies naturally represent the finer structure of the image such as fore- and background texture, and such details are not usually necessary for a successful classification (in this domain) to be made. Given this fact, only the lowest 16 spatial frequencies were used in the final stage of pre-

Table 1. Consistency of subjects (Experts A to H) in *Ceratium* classification trials

External consistency (rounded to %):								
	A	B	C	D	E	F	G	H
A	99	95	95	78	74	53	43	53
B	95	97	95	78	75	53	44	57
C	95	95	97	76	74	53	43	55
D	78	78	76	92	78	63	53	56
E	74	75	74	78	92	61	57	54
F	53	53	53	63	61	91	58	50
G	43	44	43	53	57	58	88	52
H	53	57	55	56	54	50	52	85
International inconsistency (errors per 110 pairs):								
	3	6	6	17	18	20	27	33

processing. This reduces the amount of information being used by a factor of 4, while retaining important information. It had previously been found that training a back-propagation network on the raw spatial frequency amplitudes led to the total system error's oscillating and failing to be reduced to a satisfactory level (Simpson 1990). A solution to this problem was to take the 'frequency gradient' as the data to be given to the network. This is simply the result of subtracting the power of frequency $n+1$ from the power of frequency n . Thus 16 frequencies are converted into 15 differences of frequencies – a simple gradient measure (Fig. 3d). These values were then normalised (i.e. sum to 1).

For each trial a network with 15 input, 4 hidden and 2 output nodes was set up. Each layer in the network was fully connected to the layer above (i.e. the input layer was connected to the hidden layer, and the hidden layer to the output layer). Fully connected implies that each node in one layer was connected to each node in the layer above. Training the network on a set of patterns consisted of the following: first a pattern (plus a small amount of noise) would be clamped onto the input layer of the network, so that the values of those nodes would become the values of the elements of the input pattern. This activation was then allowed to propagate through the hidden layer, to the output layer. The error at the output layer (i.e. the difference between actual output and the desired output) was then used to calculate the changes in the weights of the whole network (for an exposition of this see Rumelhart et al. 1986). Then the network was trained on another pattern from the training set, and so on until all the patterns in the training set had been used. This constituted one *epoch* of learning. Many training epochs are performed until the network is performing the classification task appropriately well, and then the network is tested on the test set of data, none of which it has been exposed to. If the network performs significantly well, generalisation is said to have occurred.

RESULTS

One-hundred simulations were carried out. For each trial, a random number between 2 and 45 inclusive was generated, and taken as the number of both *Ceratium arcticum* and *C. longipes* patterns in the training set. Hence the network was trained on between 4 and 90 patterns, with an equal number of the 2 species being presented. Each new network was initialised with random weights, and trained until the root mean squared (RMS) error (of actual outputs compared with desired outputs) reached 1% or less, or 1000 presenta-

tions of each training example had been presented, whichever occurred first. In 66% of the trials the learning was halted because the network had satisfactorily learned to classify the training data. The results of these 'successful' trials are illustrated in Fig. 4. In those cases where the learning was halted due to 1000 iterations having been completed, it was found that the network was in fact learning (i.e. the error was still decreasing), but at a very slow rate. In no cases did the network not appear to be learning the training set, and indeed in all but 8 cases (92% of the time) significant generalisation occurred. In Fig. 4 those cases when the generalisation is not considered significant (the binomial probability is >0.05) are indicated. The network is then said to have learned the training set, i.e. given a pattern from this set, the correct classification pattern (within some degree of accuracy) will be generated on the output nodes. In fact, by treating an output pattern as representing that allowable classification to which it is nearest, the output can be considered to be that classification – a nearest neighbour classifier in effect. In this way the network is 'forced' to decide between the allowed classes. Using this technique for assessing the performance, the network was then tested on the unseen test data. The results are shown in Fig. 4. The generalisation performances attained on the very small training sets are near 0.5. There is clearly a trend for a network's generalisation ability to improve with an increase in the size of training set (Pearson's product-moment correlation; for all successful trials, $r = -0.626$, $df = 64$, $p < 0.001$, and $C(-0.754 \leq \rho \leq -0.452) > 0.95$; for all trials, successful or not, $r = -0.561$, $df = 98$, $p < 0.001$, and $C(-0.683 \leq \rho \leq -0.409) > 0.95$). These trends cannot, however, be regarded as linear since it would be impossible for the network to exceed 100% performance. Since any underlying trend must therefore be non-linear, the correlation measure will underestimate the association between the 2 variables. Performance in the 'unsuccessful' trials was often as good as in the 'successful' trials: indeed, a performance of less than 10% error

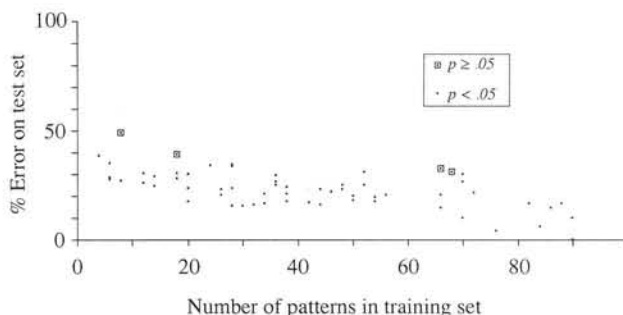


Fig. 4. Generalisation on test data vs size of training set

was attained over a test set of 30 patterns, although this is not shown in Fig. 4, the network having been halted after 1000 iterations. Even with small training sets most networks are able to demonstrate a classificatory ability beyond chance (i.e. generalising significantly). This suggests that *over-learning*, or simple memorisation of the data, is not occurring. This is probably due to the small number of hidden nodes providing a bottleneck for the input patterns, and thereby enforcing encoding on those nodes alone. This, together with the trends demonstrated above, suggests that a larger training set size than used at present will result in a network able to classify a large test set with a high degree of accuracy.

CONCLUSION

It has been shown that a neural network with 3 layers of weights is capable of learning numerous real-world data by the backward error propagation method. This technique has been used to assess the possibility of enabling a network to classify an unknown data set in the *Ceratium* domain, by training the network on a finite training set. Given the small amount of data used in this experiment, and the high proportion of significant results, it seems that such a technique has the possibility of being extended in 2 ways. Firstly, a larger data set would enable us to train a network to a much greater degree of skill, and the suggestion is that such a network could be capable of robust and accurate classification. This would not be a negligible achievement, since the task of *Ceratium* classification requires much experience in order to be performed competently. Even with such experience it remains a partly undetermined task, as evidenced by the ratings of human classifiers in Table 1. Secondly, it is thought that this technique may successfully be applied to other *Ceratium* species, such as *C. horridum* (Cleve) Gran., and *C. tripos* (O. F. Müller) Nitzsch. It appears that discriminating between these further classes and *C. longipes* and *C. arcticum* is qualitatively

a similar task, generally relying on observation of the characteristics of the apical and the 2 antapical horns. Work in this direction is in progress at Polytechnic South West in conjunction with the Alister Hardy Foundation for Ocean Science.

Acknowledgements. We gratefully acknowledge the kind assistance of Plymouth Marine Laboratory, and the analysis team of the continuous plankton recorder (Sir Alister Hardy Foundation for Ocean Science).

LITERATURE CITED

- Colebrook, J. M. (1982). Continuous plankton records: seasonal variations in the distribution and abundance of plankton in the North Atlantic Ocean and the North Sea. *J. Plankton Res.* 4: 435–463
- Glover, R. S. (1967). The continuous plankton recorder survey of the North Atlantic. *Symp. Zool. Soc. Lond.* 19: 189–210
- Jeffries, H. P., Berman, M. S., Poularikas, A. D., Katsinis, C., Melas, I., Sherman, K., Bivins, L. (1984). Automated sizing, counting and identification of zooplankton by pattern recognition. *Mar. Biol.* 78: 329–334
- Jeffries, H. P., Sherman, K., Maurer, R., Katsinis, C. (1980). Computer-processing of zooplankton samples. In: Kennedy, V. (ed.) *Estuarine perspectives*. Academic Press, New York, p. 303–316
- Rolke, M., Lenz, J. (1984). Size structure analysis of zooplankton samples by means of an automated image analyzing system. *J. Plankton Res.* 6: 637–645
- Rumelhart, D. E., Hinton, G. E., Williams, R. J. (1986). Learning internal representations by error propagation. In: Rumelhart, D., McClelland, J. (eds.) *Parallel distributed processing*, Vol. 1. MIT Press, Cambridge
- Simpson, R. G. (1989). Using Fourier analysis as pre-processing for input to a neural network model. *PSW Tech. Rep. CX3*, October
- Simpson, R. G. (1990). *Ceratium* classification in a back-propagation network. *PSW Tech. Rep. CX6*, January
- Steidinger, K. A., Chase, C., Garrett, J., Mahmoudi, B., Roberts, B., Tomas, C., Truby, E. (1990). The use of optical pattern recognition in dinoflagellate taxonomy. In: Granéli, E., Sundstrom, B., Edler, L., Anderson, D. M. (eds.) *Toxic marine phytoplankton*. Elsevier, Amsterdam
- Yarranton, G. A. (1967). Parameters for use in distinguishing populations of *Euceratium* Gran. *Bull. mar. Ecol.* 6: 147–158

This article was submitted to the editor

Manuscript first received: July 5, 1991

Revised version accepted: January 9, 1992