Vol. 139: 281-287, 1996 | MARINE ECOLOGY PROGRESS SERIES Mar Ecol Prog Ser | Published August 29

Automatic classification of field-collected dinoflagellates by artificial neural network

P. F. Culverhouse[1,*], R. G. Simpson[1], R. Ellis[2], J. A. Lindley[3], R. Williams[3], T. Parisini[4,**], B. Reguera[5], I. Bravo[5], R. Zoppoli[4], G. Earnshaw[1], H. McCall[3], G. Smith[2]

[1]Centre for Intelligent Systems, School of Electrical, Electronic and Communications Engineering, [2]Department of Psychology, University of Plymouth, Plymouth PL4 8AA, United Kingdom

[3]Plymouth Marine Laboratory, Prospect Place, West Hoe, Plymouth PL1 3DH, United Kingdom

[4]Dipartimento Informatico Systemistica Telematico, Universita di Genova, Via Opera Pia 13, I-16145 Genova, Italy

[5]Centro Oceanografico de Vigo, IEO, Apdo. 1552, E-36280 Vigo, Spain

ABSTRACT: Automatic taxonomic categorisation of 23 species of dinoflagellates was demonstrated using field-collected specimens. These dinoflagellates have been responsible for the majority of toxic and noxious phytoplankton blooms which have occurred in the coastal waters of the European Union in recent years and make severe impact on the aquaculture industry. The performance by human 'expert' ecologists/taxonomists in identifying these species was compared to that achieved by 2 artificial neural network classifiers (multilayer perceptron and radial basis function networks) and 2 other statistical techniques, k-Nearest Neighbour and Quadratic Discriminant Analysis. The neural network classifiers outperform the classical statistical techniques. Over extended trials, the human experts averaged 85% while the radial basis network achieved a best performance of 83%, the multilayer perceptron 66%, k-Nearest Neighbour 60%, and the Quadratic Discriminant Analysis 56%.

KEY WORDS: Taxonomic categorisation · Neural networks · Dinoflagellates

INTRODUCTION

With the implementation of the European Community Directive on the Quality of Water for Bivalve Cultivation, the monitoring of noxious and toxic algae and other parameters in coastal waters has become an obligation for the authorities controlling the marketing of bivalves. Monitoring programmes are very costly, especially in terms of specialised scientists and assistants who have to spend many hours at the microscope identifying and counting phytoplankton taken at weekly intervals over many stations within the European coastal zone.

Recent research has highlighted the ease of use of artificial neural networks for the visual classification task by applying the techniques to the identification of a variety of marine plankton species, including dinoflagellates (Simpson et al. 1991, 1992), 5 species of tintinnid (Culverhouse et al. 1994) and 4 species of fish larvae (Culverhouse 1995), as well as to the correlation of toxins to fish liver degradation (Ellis et al. 1994). The work reported in this paper extends these results to the development of artificial neural network classifiers for the automatic categorisation of 23 species of toxic and noxious dinoflagellate species. Dinoflagellates were selected for this exercise because of the difficulty in taxonomic discrimination of species compounded by the various morphotypes which can occur (López 1966, Bravo et al. 1995a, b, Subba Rao 1995, McCall et al. 1996, Reguera et al. 1996).

The use of automatic physical and chemical measuring systems has provided a substantial increase in the ability to monitor effectively the environmental parameters of coastal waters. There is nothing equivalent for plankton sampling except for recent developments

*E-mail: p.culverhouse@plymouth.ac.uk

**Present address: Department of Electrical, Electronic and Computer Engineering, University of Trieste, Via Valerio 10, I-34175 Trieste, Italy

of the flow cytometer for phytoplankton identification. These are still limited in their discriminatory abilities, although 20-species classification has been demonstrated for culture populations of phytoplankton (Boddy & Morris 1993).

METHODS

Field samples were used in order to reflect the natural variance in the morphology of the species and to include damaged and detritus-contaminated samples. Two artificial neural network (ANN) architectures were studied, the Radial Basis Function (RBF) classifier (Broomhead & Lowe 1988, Caiti et al. 1994) and the multi-layer perceptron classifier, back propagation of error variant (BPN) (Rumelhart et al. 1986). The performance and behaviours of these 2 novel classifiers were compared to 2 classical multivariate statistical techniques, k-Nearest Neighbour (kNN) and Quadratic Discriminant Analysis (QDA) (Kendall 1966).

The microscopic images were pre-processed to segment the specimen from the background, debris and clutter in the image. This was carried out automatically, a process which resulted in images holding only the specimen of interest, but including any debris and clutter in contact with the specimens (Simpson et al. 1991, 1992, McCall et al. 1996). This additional material could be considered as noise to the recognition process but did not prevent the neural networks from functioning correctly. Two types of image figure-ground separation were tested, one based upon a Sobel edge density gradient, the other on the distributions of Gabor set derived textures (Gabor 1946, Daugman 1990). Sobel edge density gradient was used in this study. The isolated specimen images were then analysed by 6 functions to derive a multiplicity of low resolution parameters with which to feed the automatic classifiers (Gonzalez & Woods 1992, Ellis et al. 1994). For example, only the first 15 of 128 frequency bins resulting from a 2D Fourier transform analysis were saved (Simpson et al. 1991). The functions were the 2D Fast Fourier Transform of the object, the Discrete Fourier Transform of the object's profile, its second order statistics, a Sobel edge descriptor, a junction descriptor and a texture metric. Functions were chosen to provide non-overlapping partial measures of an object's shape and surface texture. The data set thus comprised a collection of 60 variables describing each of the specimens in the pool for training and testing the automatic classifiers (Ellis et al. 1994).

Coarse coding is an attempt to model behavioural features of the mammalian visual system. Humans can learn to perform almost arbitrary visual discriminations given sufficient training. It has been suggested that our visual system is made up of a large set of basic, general purpose visual operations or routines from which a sub-set may be composed, during training, for the purposes of specific visual tasks (Ullman 1984). Since it is supposed that arbitrary sequences of these routines can be composed, they offer an explanation for both universal human visual abilities, such as recognising the faces of our family, and expert abilities, such as being able to sort species of *Ceratium*. This model of human visual abilities inspires our proposals for general purpose network classifiers which have the following features. Input units will be divided into a (large) set of channels. Each channel will carry a different class of visual information. The information will, typically, be of low resolution, so as to escape the notorious difficulty of obtaining high precision visual information from images.

Classifier training consists of selecting 100 random sets of specimen data from the data pool described above, presenting each classifier with the data and allowing it to settle to a stable state, thereby forming the model (Fig. 1). Unused data from the data pool is then used to test the classifier's performance on independent material. The assumption is that the archive of field-collected specimens is a uniformly distributed sub-set of the natural population. Therefore random selection of data for training and testing is held to be representative of the natural populations and the categorisation results obtained from the automatic classifiers are also representative of their behaviours on fresh specimens collected from the field. We consider this to be true because the periods of collection of specimens during the project were spread over a 3 yr work programme. Specimens were obtained from the Plymouth Marine Laboratory (UK) plankton archive as well as from fresh-collected material from the Centro Oceanografia de Vigo (Spain) and thus represented collections from differing geographical areas of the eastern seaboard of the European Union and the North Atlantic Ocean. Sampling localities were not correlated against morphological variances, as the purpose of this study was to explore the ability of neural networks to categorise all the morphotypes of a particular species as one species. Another study, not reported here, reviewed the neural network processing required to make these morphological distinctions.

The training and test protocols were normally repeated many times to gain a mean performance across many random data pools.

Over 5000 field-collected specimens were identified, photographed and labelled by a team of expert taxonomists/ecologists, according to a multi-criterion questionnaire. The resulting database, known as the mas-

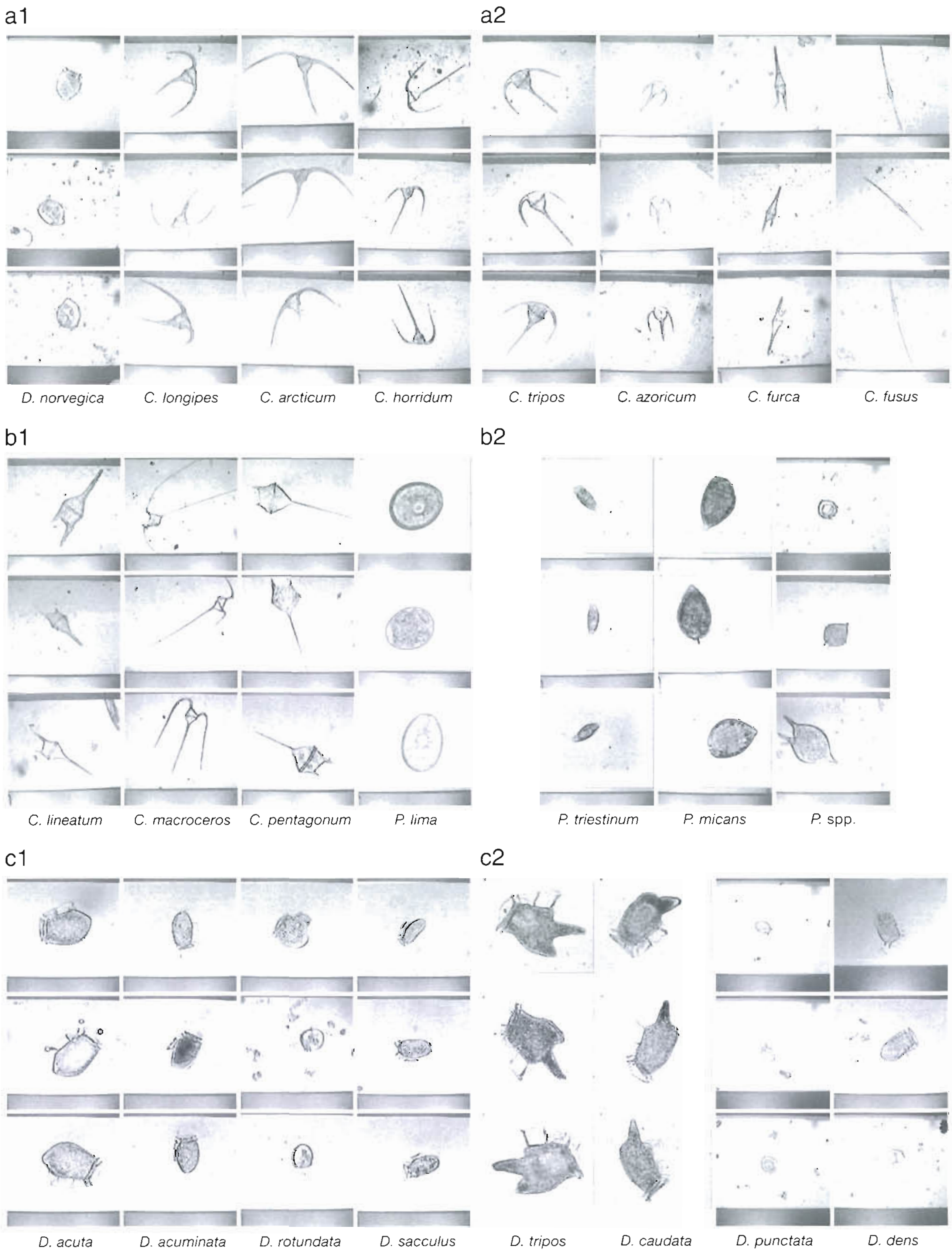

Fig. 1. Three examples of 23-species dinoflagellate data sets

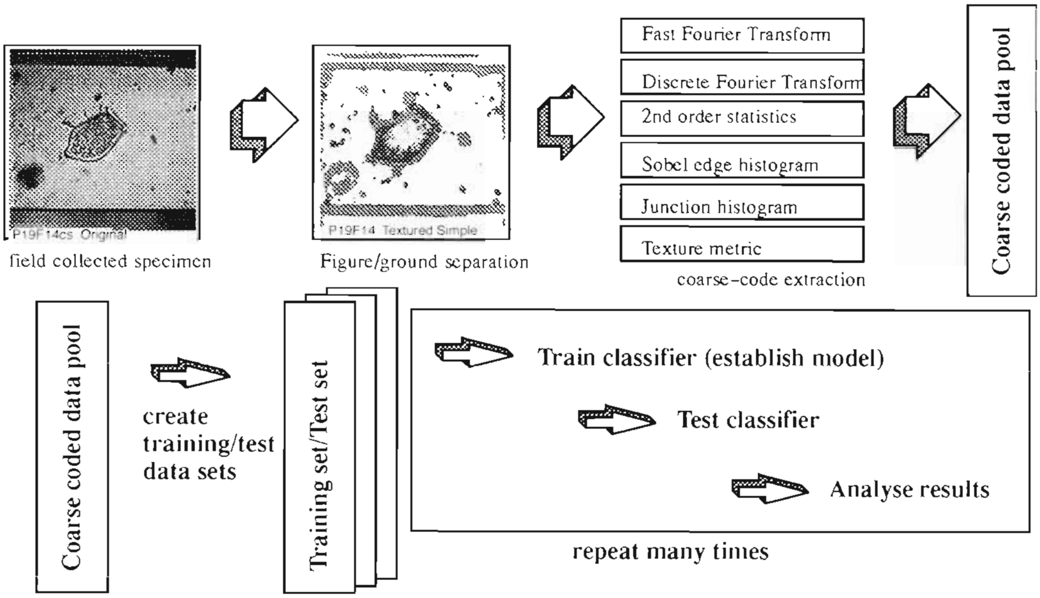

Fig. 2. Outline of experimental protocol employed in ANN experiments

terlog, was used to select specimens according to the certainty of their taxonomic label and their image quality (including level of debris and clutter). Photomicrographs of each specimen were archived. Computer-digitised images of these photomicrographs were required to allow computer-based neural network studies. These were also archived. Data suitable for numerical analysis and neural network studies were extracted from the selected images (Fig. 2).

It was noticed during early trials that the results were biased toward particular photomicrograph film stock, a so-called film complicity to classification. Additional constraints were therefore placed on the selection of data for training regimes, to ensure no bias remained. This bias is a problem that may be resolved by modifying the specimen image capture methodology from indirect sampling (computer digitise each photomicrograph) to direct sampling (computer digitise specimen under microscope), and by calibrating each microscope employed to a reference.

RESULTS

Categoriser performance

A series of extended studies of ANN and statistical categorisers were undertaken. Fig. 3 summarises the performance across dinoflagellate species. There was no statistical difference between the RBF and BPN classifier results over 14 or 23 species. The differences in performance between 4 species (*Dinophysis acuta, D. acuminata, D. rotundata, D. sacculus*), 5 species (*Ceratium arcticum, C. azoricum, C. horridum, C. longipes, C. tripos)* and 9 species (*C. arcticum, C. azoricum, C. horridum, C. longipes, C. tripos, D. acuta, D. acuminata, D. rotundata, D. sacculus)* trials (Simpson et al. 1994, Ellis et al. 1996) and 14 species trials are as a result of changing the neural network pre-processing regime. The coarse coding concept was introduced to the 14 species (*C. arcticum, C. azoricum, C. furca, C. fusus, C. horridum, C. lineatum, C. longipes, C. macroceros, C. pentagonum, C. tripos, D. acuta, D. acuminata, D. rotundata, D. sacculus)* trials, but not the earlier trials. This change also allows a graceful expansion to the system, by the simple act of adding more coarse coding channels of pre-processing to the system (see Fig 2).

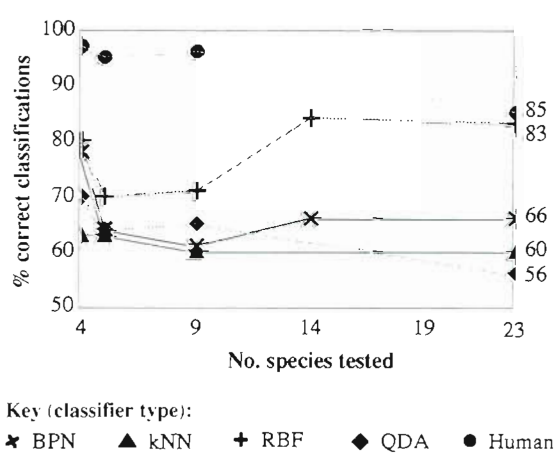

Fig. 3. Summary of classifier performance over dinoflagellate species

Expert performance was ascertained by random presentation of specimen images to individuals comprising a panel of competent taxonomists. Identification accuracies were in the range 95 to 97% over 2 to 9 species problems. Results from a study of human performance over 23 species indicate an increase in error rate as the number of categories increases, giving a performance of between 83 and 86%. Comparing ANN to expert performance, experts are clearly competent at taxonomy, but their performance is affected by a number of psychological factors (Chomsky 1972), including the human short-term memory limit of 5 to 9 items stored, fatigue and boredom (Colquhoun 1971, Davies & Parasuraman 1982), recency effects (where a new classification is biased toward those in the set of most recently used labels) and positivity bias (where labelling a specimen is biased by one's expectations of the species present in the sample). In contrast ANN expert competence is fully reflected in their performance.

Automatic classifier performances are similar in behaviour over the range of species, with the RBF neural network leading with an accuracy of 83% best performance over 23 species (*Ceratium arcticum, C. azoricum, C. furca, C. fusus, C. horridum, C. lineatum, C. longipes, C. macroceros, C. pentagonum, C. tripos, Dinophysis acuta, D. acuminata, D. caudata, D. dens, D. norvegica, D. punctata, D. rotundata, D. sacculus, D. tripos, Prorocentrum lima, P. micans, P. triestinum, Peridinium* spp.) labelling tasks; BPN follows with 66% accuracy. Both the multivariate statistic models (kNN and QDA) lag with 60% and 56% performance respectively.

Table 1 shows the detailed performance across the 4 classifiers by species for data set *bmix2*, one of the randomised data sets drawn from the data archive. It may be seen that the performance of all classifiers varies with particular species. It appears that this behaviour is correlated to population size and to population morphological variance, as shown for a sub-set of these data in McCall et al. (1996). Large intra-species data pools with low morphological variances within the pools give rise to high accuracies of identification.

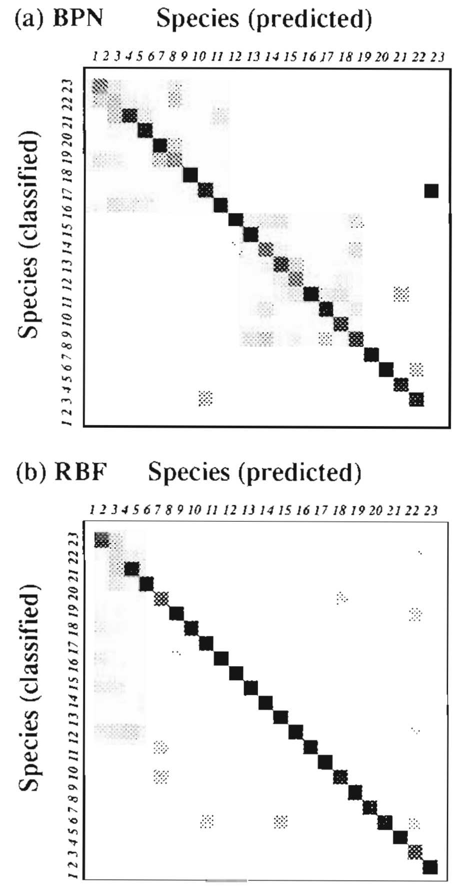

Fig. 4. Confusion table plots for the 23-species data set (see Table 1 for key)

Fig. 4 illustrates the confusion tables for the networks graphically; a black diagonal line indicates no confusions, any deviations from this highlights misclassifications. The levels of grey indicate the certainty of classification with black being very certain and light grey being very uncertain. In Fig. 4a it is noticeable that there are more confusions between species of *Dinophysis* than there are between *Dinophysis* and other genera (elements of top left corner). In contrast, Fig 4b shows that the mis-classifications of the RBF network are more uniform across genera. This is

Table 1. Classification performance (%) for edited 23-species data (rounded to integer) for data set *bmix2*. Key — Species 1: *Dinophysis acuta*; 2: *D. acuminata*; 3: *D. rotundata*; 4: *D. sacculus*; 5: *Ceratium longipes*; 6: *C. arcticum*; 7: *C. horridum*; 8: *C. tripos*; 9: *C. azoricum*; 10: *C. furca*; 11: *C. fusus*; 12: *C. lineatum*; 13: *C. macroceros*; 14: *C. pentagonum*; 15: *Prorocentrum lima*; 16: *P. triestinum*; 17: *P. micans*; 18: *Dinophysis tripos*; 19: *D. caudata*; 20: *D. punctata*; 21: *D. dens*; 22: *D. norvegica*; 23: *Peridinium* spp.

Spp.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	Mean
BPN	79	56	64	77	53	81	49	60	75	85	67	55	27	67	86	67	67	100	63	0	40	82	0	**61**
RBF	81	58	73	82	62	94	68	76	75	85	81	79	90	89	86	100	100	100	85	100	68	89	100	**83**
QDA	87	52	59	75	53	57	42	84	89	36	67	30	45	79	0	0	100	67	68	100	0	64	0	**60**
kNN	72	50	73	68	40	84	45	43	7	75	88	33	50	70	29	67	67	100	45	22	0	67	0	**52**

expected since the RBF network made only 17% incorrect classifications as compared to the BPN 34% error rate.

DISCUSSION

Four automatic categorisation algorithms have been evaluated. The performance of the best algorithm approaches that of human experts on labelling experiments using 23 species of dinoflagellate (see Fig 3). This is an extremely important result since the images employed in the studies were of field-collected specimens, not from culture collections. As a result they exhibited both morphological variance and variable image quality. In particular many of the *Dinophysis* spp. and *Ceratium* spp. showed considerable intraspecies morphological variance, with many specimens requiring an expert's opinion to complete the classification of the species within the samples. Both detritus and other specimens often cluttered the field of view; sometimes direct contact with the specimen resulted in the debris being incorporated into the processing scheme, and so participated in the classifier performance evaluations.

Normal practice in image processing would dictate that the 3 microscopes used in the data collection at the various sites were precisely calibrated and that a uniform specimen handling protocol was observed. These procedures would ensure that the data were both corrected for geometric distortions of scale and for any different staining and preserving procedures. We did not attempt to make these corrections, and thus operated under more difficult conditions as a result. This was a controlled attempt to design and produce an automatic classifying system that was resilient to these factors, enabling the classifiers to be directly applied in the field.

However, at present such a system is not easy to use. Firstly the training protocols require that each individual specimen is given an accurate taxonomic label. We have established a 2-expert protocol for this process, with taxonomically dubious labels being referred to a panel of judges. Secondly the completion of a useful system requires the collection of approximately 100 examples of what may be on occasions rare species. This target may also be further complicated by the diverse life cycles of some of the dinoflagellates such as *Dinophysis acuminata*, *D. acuta* and *D. norvegica* which show considerable variation in their morphology due to environmental conditions and life cycle stages.

Images with large amounts of detritus and multiple specimens within the field of view are common. The existing system can handle multiple specimens provided they are not overlapping; detritus contacting the specimen does not appear to affect neural network classifier performance provided the material is less than 20% of the specimen's size.

CONCLUSIONS

In extended trials the artificial neural network classifier systems gave better performances than the multivariate statistical systems. There is no basis as yet for describing the differences in detail, but if neural networks are conceived of as not merely statistical systems but as representational systems, then their greater power may be explained through their ability to form complex encodings of the categories. These encodings may capture regularities which are not expressable in any of the existing standard statistical descriptions.

A uniform pre-processing method has been developed which operates on images that may be cluttered and partially filled with debris. Although this may have resulted in specimens which were morphologically modified by attached debris, the automatic classifiers could still recognise them. This inexact figure-ground separation simplifies the task of the image pre-processing stages of machine classifying systems.

Developments in texture pre-processing should extend the pre-processing to allow classifiers to operate on highly cluttered and debris-filled images. Research on visual object grouping mechanisms in human subjects and in artificial neural networks indicates that extensions to the categorisers are possible, which will enable them to deal with highly cluttered images, and images with overlapping specimens.

The system developed has successfully been applied to specimens of Dinophyceae, Tintinnidae and fish larvae (Culverhouse et al. 1994, Culverhouse 1995). The performances of the neural networks, when tested on previously unseen data, compares favourably with human classification studies of the specimens, and also with 2 classical statistical clustering methods.

A series of categorisation experiments carried out over a number of different species populations indicates that the classifier protocols and structures developed may well scale up to much larger systems without modification. It is anticipated that refinements to the coarse coded parameter extraction methods will take ANN performance to above 90%.

Automatic categorising systems such as these are inexpensive to implement as assistant taxonomic categorisers, requiring only a microscope, a personal computer, television camera interface and software. This makes them commercially attractive, and will perhaps enable these routine ecological and biological assays to be completed automatically. The potential of such systems to reduce sample analysis time is enormous. For example, a routine plankton sample, if done manually, may take 2 h to analyse and to log the resulting data in a spreadsheet. We predict that this could be reduced to 5 min with computer-based neural network categorisation systems.

Acknowledgement. The authors gratefully acknowledge the support of the European Union given under MAST 2 programme contract CT92-0015.

LITERATURE CITED

Boddy L, Morris CW (1993) Analysis of flow cytometry data — a neural network approach. Binary 5:17–22

Bravo I, Deladago M, Fraga S, Honsell G, Lassus P, Montresor M, Sampayo MA (1995a) The *Dinophysis* genus: toxicity and species definition in Europe. In: Lassus P, Arzul G, Erard-le Denn E, Gentien P, Marcaillou-le Baut C (eds) Toxic phytoplankton blooms in the sea. Lavoisier Publishing, Paris, p 843–845

Bravo I, Reguera B, Fraga S (1995b) Description of different morphotypes of *D. acuminata* complex in the Galician rias Bajas in 1991. In: Lassus P, Arzul G, Erard-le Denn E, Gentien P, Marcaillou-le Baut C (eds) Toxic phytoplankton blooms in the sea. Lavoisier Publishing, Paris, p 21–26

Broomhead DS, Lowe D (1988) Multivariable functional interpolation and adaptive networks. Complex Systems 2: 321–355

Caiti A, Magenes G, Parisini T, Simpson R (1994) Smooth approximation by RBFs: three case studies. J Appl Sci Comp 1:88–112

Chomsky N (1972) Language and mind. Harcourt, New York

Colquhoun WP (ed) (1971) Biological rhythms and human performance. Academic Press, London

Culverhouse PF (1995) Automatic methods for the detection of harmful algae. In: Lassus P, Arzul G, Erard-le Denn E, Gentien P, Marcaillou-le Baut C (eds) Toxic phytoplankton blooms in the sea. Lavoisier Publishing, Paris, p 695–706

Culverhouse PF, Ellis R, Simpson RG, Williams R, Pierce RW, Turner JT (1994) Categorisation of 5 species of *Cymatocylis* (Tintinidae) by artificial neural network. Mar Ecol Prog Ser 107:273–280

Daugman JG (1990) An information-theoretic view of analogue representation in striate cortex. In: Schwartz EL (ed) Computational neuroscience. MIT Press, Cambridge, MA, p 403–424

Davies DR, Parasuraman R (1982) The psychology of vigilance. Academic Press, London

Ellis R, Simpson RG, Culverhouse PF, Parisini T (1996) Committees, collectives and individuals: expert visual classification by neural network. Neural Comp Appl (in press)

Ellis R, Simpson RG, Culverhouse PF, Parisini T, Williams R, Reguera B, Moore B, Lowe D (1994) Expert visual classification and neural networks: can general solutions be found? In: IEEE Oceans '94, Brest, September 1994, Piscataway, NJ, p 330–334

Gabor D (1946) Theory of communication. J Inst Elect Eng 93:429–457

Gonzalez RC, Woods RE (1992) Digital image processing. Addison-Wesley, New York

Kendall DG (1966) Discrimination and classification. In: Krishnaiah PR (ed) Multivariate analysis, Vol 1. Academic Press, New York

López J (1966) Variación y regulación de la forma en el género *Ceratium*. Invest Pesq 30:327–427

McCall H, Bravo I, Lindley JA, Reguera B (1996) Phytoplankton recognition using parameteric discriminants. J Plankton Res 18(3):393–410

Reguera B, Bravo I, McCall H, Reyero MI (1996) Phased cell division and other biological observations in field populations of *Dinophysis* during cell cycle studies. In: 7th Int Conf Toxic Marine Phytoplankton, Sendai, Japan, July 1995. IOC, UNESCO (in press)

Rumelhart DE, Hinton GE, Williams RJ (1986) In: Rumelhart DE, McClelland J (eds) Parallel distributed processing, Vol 1. MIT Press, Cambridge, MA, p 318–362

Simpson R, Culverhouse PF, Ellis R, Williams R (1991) Classification of *Euceratium* Gran. in neural networks. In: IEEE Int Conf on Neural Networks in Ocean Engineering, Washington DC, USA, August, p 223–230

Simpson R, Culverhouse PF, Ellis R, Williams R (1994) Classification of *Ceratium* and Dinophyceae plankton species. In: Marinaro M, Morasso PG (eds) Proc Int Conf on Artificial Neural Networks, ICANN '94, Sorrento, Italy, 26–29 May 1994. Springer-Verlag, Berlin, p 843–846

Simpson R, Williams R, Ellis R, Culverhouse PF (1992) Biological pattern recognition by neural networks. Mar Ecol Prog Ser 79:303–308

Subba Rao DV (1995) Life cycle and reproduction of the dinoflagellate *Dinophysis novegica*. Aquat Microb Ecol 9:199–201

Ullman S (1984) Visual routines. Cognition 18:97–159

This article was submitted to the editor

Manuscript first received: January 29, 1996
Revised version accepted: April 16, 1996