Transactions on Information and Communications Technologies vol 1, © 1993 WIT Press, www.witpress.com, ISSN 1743-3517

Biological classification of river water quality using neural networks

B.M. Ruck, W.J. Walley & H.A. Hawkes
Department of Civil Engineering, Aston University, Aston Triangle, Birmingham B4 7ET, UK

ABSTRACT

The paper describes an investigation into the use of neural networks for the direct classification of river water quality from biological data. The theoretical basis of biological monitoring is briefly explained and the most commonly used methods are discussed, together with some recently developed computer-based methods. The biological basis of the study is fully described, while it is assumed that the reader has a basic understanding of artificial neural networks. The networks used were multi-layer perceptrons with a single hidden layer of eight nodes and an output layer of five nodes, one for each of the five biological-based water quality classes adopted for use in this study. Two different input sets were tested: one based upon the states of existence, in field samples, of forty-one key biological indicator organisms; and the other based upon twelve inputs derived from principal component analysis of the field data. Fifty-three field samples, previously classified by an expert river ecologist, were used for both the training and the testing of the networks, but independence between the training and test sets was maintained using one-fold cross validation. Despite the limitations of the data set, the results show that neural networks have potential for use as direct classifiers of river water quality from biological field data.

INTRODUCTION

Increasing importance is being placed on the use of riverine ecology as a means of monitoring and classifying river quality, both in terms of its water quality and its broader environmental quality. The various biological flora and fauna, such as attached algae, macrophytes and benthic (or river bed) macro-invertebrates, are seen as continuous monitors of the river's 'health', and field data on these are used to classify the river. In the case of water quality monitoring and classification, the biological methods are used to complement the more traditional chemical methods. This paper describes a new approach to the interpretation of biological data, based on neural network techniques. It is applied to the classification of river water quality, but could equally be applied to the classification of environmental quality. For a fuller introduction to biological monitoring theory and techniques the interested reader is referred to De Pauw and Hawkes [5], Hawkes [6,7,8,9], Hawkes and Hughes [10], Hellawell [11] and Metcalfe [15].

At present the most suitable single group for monitoring purposes is considered to be the benthic macro-invertebrates. These animals form part of the community associated with the river bed and are relatively immobile, thus they are representative of the sample location. They are present in all rivers, except in cases of extreme pollution, and cover a range of life modes and trophic levels. Taxonomic classification of these organisms to family or genus level is not exacting, and qualitative identification can often be carried out at the river bank. Also, the different species are known to have different sensitivities to pollutants, thus the structure of the benthic macro-invertebrate community is affected by both degradable organic matter (sewage) and toxic pollutants (pesticides and heavy metals).

Present U.K. methods of interpreting samples of benthic macro-invertebrates are based around score systems. The most commonly used systems in Europe are based on the British Monitoring Working Party (BMWP) score (Department of Environment [4]), the Average Score Per Taxon (ASPT), the Trent Biotic Index (TBI) (Woodiwiss [23]) and the Saprobic Index (Sladecek [20]). The BMWP system allocates a number to each family that is indicative of its sensitivity to organic pollution. The overall BMWP score is then calculated from the summation of these numbers for all the families present in the sample, and the ASPT is derived by dividing the BMWP score by the number of scoring families present. The TBI uses the most sensitive taxa present to determine the base point on a classification table and then uses the total number of groups present to pinpoint the final score. The BMWP and TBI systems do not take into account the abundance of the taxa, only their presence or absence, but the Saprobic Index and Chandler score (Chandler [3]) do incorporate abundance. In addition, the Saprobic Index, being based upon the Saprobic System of Kolkwitz and Marsson [13, 14], accommodates flora and fauna other than the benthic macro-invertebrates. All of these systems are subjective since there is no fundamental basis for the numbers they used. Consequently, widely different view points are held by biologists on their relative merits.

A recent development is the River Invertebrate Prediction and Classification System (RIVPACS) (Moss *et. al.* [17], Wright *et. al.* [24]), which uses multi-variate statistical techniques to predict the benthic community structure at any site in the UK on the assumption that the river is unpolluted . From this structure the expected number of taxa, BMWP score and ASPT are determined and then compared with their actual values derived from samples taken at the site. Ratios of the actual to expected values, termed Environmental Quality Indices (EQI), are then used to represent the environmental stress at the site. The National Rivers Authority [18] has proposed that this be incorporated into a new classification system for the UK. In its present form however the RIVPACS system is strictly only suitable for the classification of environmental quality, not water quality. To be an effective classifier of water quality its development process would have to be repeated using samples taken from a single biotope, preferably riffles.

More recently, a team of researchers at Aston University has been investigating the possibility of applying some of the techniques of Artificial Intelligence (AI) to river water quality monitoring. In particular, they have been examining the potential of two very different approaches (Walley [21]): a knowledge-based systems approach based on methods of reasoning under uncertainty; and a neural networks approach. The former has already produced some encouraging results (Walley *et al.* [22], Boyd *et al.* [2]), and the first results of the latter are presented here in this paper. The paper assumes that the reader has a basic knowledge of neural networks, but for those who do not the authors recommend the introductory text by Hertz *et al.* [12].

BIOLOGICAL FOUNDATIONS OF THE STUDY

The aim of this project was to develop a neural network capable of directly interpreting benthic samples of macro-invertebrate into water quality terms. The water quality classes used were designed to mirror the five chemical classes (1a, 1b, 2, 3 and 4) presently in use in the UK, and were designated B1a, B1b, B2, B3 and B4 to distinguish them from the chemical classes. Forty-one taxa were used to provide the input to the network. These were initially selected by the team's domain expert, H. A. Hawkes - a leading biologist within this field, for use in the knowledge-based systems as key indicators of river water quality. All were commonly occurring benthic macro-invertebrates, having sensitivities which collectively span the whole range of water quality. Different levels of taxonomy are used for each taxon, some being identified to species level, like *Gammarus pulex* and *Asellus aquaticus*, while others are only identified to genus or family level. All were selected for their relative ease of identification. The full list is given in Table 1.

The network attempts to provide a direct mapping from benthic sample data to river water quality class in a form which mirrors the chemical classification. This system and the knowledge-based systems mentioned earlier are the only systems which aim to do this. All other biological systems produce scores or indices which then have to be interpreted into water quality terms. In addition, few of these systems make any allowance in their calculations for the abundance levels of individual taxa in the sample.

Table 1. List of Forty-One Taxa used as Indicator Species and Definitions of Abundance Levels.

(Ranges: Rare = 1 to n_1-1; Established = n_1 to n_2-1; Abundant >= n_2.)

Taxon	n1	n2	Taxon	n1	n2
Polycelis nigra	2	10	*Ephemerella ignita*	3	20
Dendrocoelum lacteum	2	10	*Caenis spp.*	3	20
Potamopyrgus jenkinsi	3	50	*Amphinemura sulcicollis*	2	10
Bithynia tentaculata	3	20	*Leuctra spp.*	3	20
Lymnaea peregra	3	50	*Isoperla grammatica*	2	10
Planorbis spp.	2	10	*HALIPIDAE*	3	20
Ancylus fluviatilis	3	20	*DYTISCIDAE*	2	10
Sphaerium spp.	3	20	*ELMINTHIDAE*	2	10
Pisidium spp.	3	20	*Sialis lutaria*	2	10
TUBIFICIDAE	5	200	*Rhyacophila spp.*	3	20
LUMBRICULIDAE	5	100	*Glossosoma spp.*	3	50
Glossiphonia spp.	2	10	*Agapetus spp.*	3	50
Hellobdella stagnalis	2	10	*POLYCENTROPODIDAE*	3	20
Erpobdella octoculata	3	20	*Hydropsyche angustipennis*	3	50
HYDRACARINA	3	20	*Other HYDROPSYCHIDAE*	3	20
Asellus aquaticus	3	50	*HYDROPTILIDAE*	3	50
Gammarus pulex	3	50	*LIMNEPHILIDAE*	3	20
Baetis rhodani	3	50	*CERATOPOGONIDAE*	2	10
Rhithrogena spp.	3	20	*Chironomus riparius*	5	100
Heptagenia spp.	2	10	*Simulium ornatum*	3	50
Ecdyonurus spp.	3	20			

In a system based on neural networks it is possible to input the actual number of individuals found in each taxon, but in practice abundance levels are recorded on a banded scale because it is too time consuming to count the exact number of each. For example, abundance levels may be classified as rare (1-2 individuals present), common (3-10) , established (11-20), abundant (21-100), very abundant (100 +). In this study four levels of abundance were used: absent, rare, established and abundant, these being the same as those used by Walley *et. al.* [22] and Boyd *et al.* [12] in the development of their knowledge based systems. An interesting feature of this system is that the numerical values used to separate the different levels vary from taxon to taxon. Thus, for example, the range over which Tubificidae (sludge worms) are considered to be established is far greater than that for the less numerous species, like the glossiphonid family of leeches. The ranges for each taxon were selected by the domain expert in recognition of their normal relative abundances. Details of the intervals used for each of the forty-one taxa are given in Table 1.

Transactions on Information and Communications Technologies vol 1, © 1993 WIT Press, www.witpress.com, ISSN 1743-3517

When classifying the 53 field samples the expert refined the B1a to B4 scale into four sub-classes between each of the main classes. Thus a progressive improvement in water quality from B3 to B2 would on his scale be represented as B3, B3+, B3++, B2--, B2- B2, where a '+' represents an improvement and a '-' represents a degradation of the quality class. Thus each of the expert's sub-divisions represented 0.2 of a main class interval.

IMPLEMENTATION AND TESTING

The neural networks used for this study were multi-layer perceptrons, which were trained using the standard back propagation algorithm (Rumelhart *et al.* [19]). All networks had one hidden layer with 8 processing elements and an output layer with five processing elements, each corresponding to one of the five quality classes, as indicated in Figure 1.

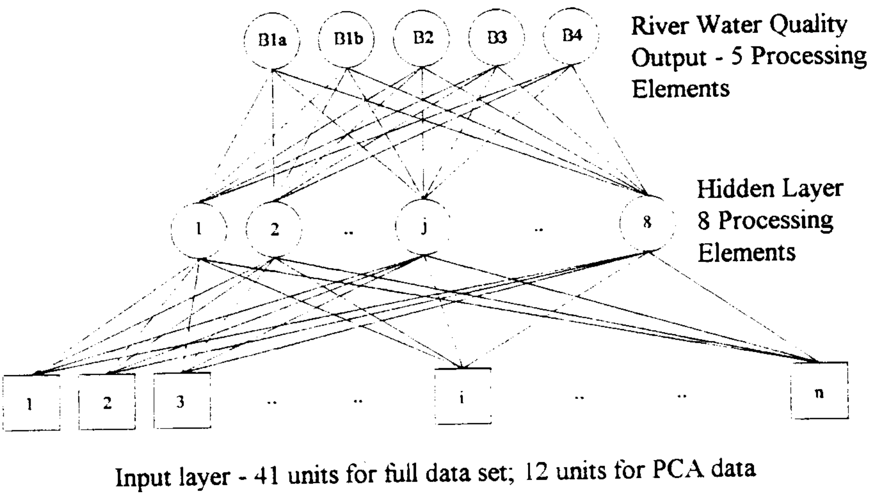

Figure 1. The network topology used in this paper.

Two different approaches to data input were tried, one using all forty-one taxa as input to the networks and the other using twelve inputs derived from the eigenvectors of a principal component analysis (PCA). The latter can be considered as a pre-processing measure designed to reduce the noise in the data and also the effect of over-parameterisation of the network.

Two different indicators were used to test the performance of the networks:

a) *The percentage of correctly predicted whole-class classifications.*
In this case the network's classification was taken as that indicated by the output node with highest activation level. .

Transactions on Information and Communications Technologies vol 1, © 1993 WIT Press, www.witpress.com, ISSN 1743-3517

b) *The root mean square of the error between the network's and expert's classification indices.*
The classification index (CI) is based on an interval scale varying from 0 for Class B1a to 4 for Class B4. Since each of the expert's sub-division represented 0.2 on this scale, Class B2+ is equivalent to a CI of 1.8 and B1b- is equivalent to 1.2. The Classification Index was first defined by Walley *et al.* [22] for the interpretation of the probabilistic output from their Bayesian model into a single classification number. Thus, in keeping with their definition, the networks' classification indices were calculated from the normalised outputs of the five output nodes, as given by equation (1) below:

$$CI_p^N = \sum_{i=1}^{53}(i-1)Y_{pi}^N \tag{1}$$

where CI_p^N is the network's Classification Index for the *p*th pattern; and

Y_{pi}^N is the normalised activation of the *i*th output for the *p*th pattern.

The root mean square of the errors (*Erms*) between the network's classification CI_p^N and the expert's classification CI_p^E is then given by:

$$E_{rms} = \left\{\frac{1}{53}\sum_{p=1}^{53}(CI_p^E - CI_p^N)^2\right\}^{\frac{1}{2}} \tag{2}$$

In view of the relatively limited amount of data available for the training of the networks, a 1-fold cross validation scheme was used to test the performance of the networks. This allows best use to be made of the available data whilst maintaining independence between the training and the test sets, thus ensuring that the tests were carried out on 'unseen' data. Cross-validation uses the "leave-one-out philosophy" of the non-parametric jack-knife technique and uses the left-out sample to test the estimator. Thus each network was trained using 52 of the 53 samples and its predictive ability then tested against the remaining sample. This process was then repeated for each and every sample.

The sigmoid activation function used within the processing elements, with the inputs and outputs mapped between 0.1 and 0.9. The input values used to represent the various states of existence, or abundance levels, of the taxa are shown in Table 2. These were selected in the light of experience gained from the expert systems studies and after carrying out a few pilot tests.

Table 2. Input Values used for the Abundance Levels.

State	Input Value
Abundant	0.9
Established	0.6
Rare	0.15
Absent	0.1

Transactions on Information and Communications Technologies vol 1, © 1993 WIT Press, www.witpress.com, ISSN 1743-3517

RESULTS

Table 3 below shows the classification indices produced by the networks for each of the fifty-three field samples set against the expert's classifications, also expressed in CI terms. These are shown in graphical form in Figures 2a and 2b overleaf. From the statistics shown on the graphs it is clear that the 41 raw data input network performed better than the PCA-12 input network, the former giving an E_{rms} value of 0.43 of a class interval compared to 0.52 for the latter. Table 4 shows the confusion matrix based on the results of the whole-class classification tests, and on the basis of the %Correct performance indicator this also shows that the 41 input net performed better than the 12 input net, albeit very slightly (i.e. 67.9% compared to 66.0%).

Table 3. Results of performance tests showing the a networks predicted CI and the expert classification.

Sample No.	Expert's CI^E	41 Input Network's CI	12 Input Network's CI	Sample No.	Expert's CI^E	41 Input Network's CI	12 Input Network's CI
1	1.6	1.17	1.02	28	0.0	0.37	0.04
2	0.8	1.93	2.00	29	1.6	1.52	1.09
3	1.8	1.71	1.46	30	4.0	4.00	3.59
4	0.8	1.01	1.05	31	0.8	0.40	0.96
5	1.8	1.99	2.99	32	1.2	2.20	3.00
6	0.4	0.23	0.40	33	2.6	3.00	3.00
7	1.0	0.98	0.97	34	3.0	2.92	2.97
8	2.4	3.00	2.97	35	3.0	3.35	3.06
9	2.8	2.80	3.00	36	2.8	2.96	3.00
10	2.6	2.98	3.00	37	2.8	2.72	3.00
11	2.8	2.92	3.00	38	2.8	2.42	3.04
12	2.8	3.00	3.00	39	4.0	3.95	3.42
13	3.0	2.97	3.01	40	4.0	3.36	3.15
14	2.8	2.77	2.89	41	0.8	0.95	1.00
15	2.6	2.34	2.99	42	0.8	1.04	1.12
16	1.8	1.74	1.72	43	1.8	2.68	2.41
17	2.0	2.00	2.00	44	0.2	0.10	0.50
18	4.0	3.97	3.90	45	1.8	2.99	2.77
19	3.8	3.74	3.48	46	3.8	3.98	3.99
20	2.8	3.00	3.11	47	3.6	3.61	3.98
21	4.0	4.00	3.78	48	3.0	2.89	3.61
22	3.2	3.99	3.99	49	3.0	2.99	2.99
23	3.4	3.98	3.73	50	2.6	3.00	2.99
24	3.0	2.90	2.94	51	1.0	1.52	1.96
25	0.2	0.76	0.33	52	1.8	2.84	2.67
26	0.4	0.21	0.27	53	2.8	2.15	2.13
27	0.2	0.04	0.09				

Transactions on Information and Communications Technologies vol 1, © 1993 WIT Press, www.witpress.com, ISSN 1743-3517

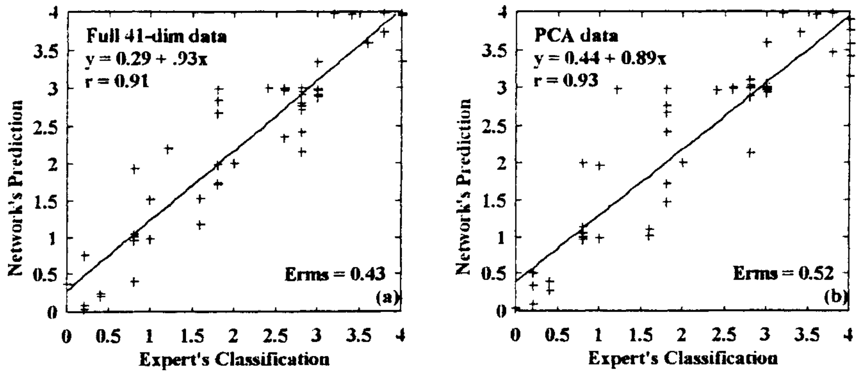

Figure 2. Results of the performance tests showing the networks' classifications plotted against the expert's for: (a) the network with inputs from all 41 taxa; and (b) the network with the 12 inputs derived from principal component analysis.

Table 4. Confusion matrices showing the performance of the two individual networks predicted classification and the expert's predicted class.

Predicted Class (41 Input Net)

Expert's Class	B1a	B1b	B2	B3	B4
B1a	5	1			
B1b	1	4	2	1	
B2		2	4	4	
B3			3	16	2
B4				1	7

% Correct = 67.9

Predicted Class (12 Input Net)

Expert's Class	B1a	B1b	B2	B3	B4
B1a	6				
B1b		5	2	1	
B2		4	2	4	
B3			1	17	3
B4				3	5

% Correct = 66.0

DISCUSSION

The results of the performance tests showed that about 67 percent of the neural networks' classifications conformed with those of the expert on a whole-class basis, and that 98 percent were within one class interval of the expert's classification. The networks which used the 12 PCA inputs performed marginally worse that those based upon the 41 raw data inputs, showing that little information, apart from noise, was lost from the dimensionality reduction. Examination of Figures 2a and 2b reveals that both networks tend to underestimate the quality of good quality waters and slightly overestimate the quality of poor quality waters. This was also found to be the case by Walley *et*

Transactions on Information and Communications Technologies vol 1, © 1993 WIT Press, www.witpress.com, ISSN 1743-3517

al.[22] during the early stages of the development of the Bayesian inference model, until it was realised that the importance of 'absence' evidence was being under-valued. With this in mind, further work is being undertaken to establish a better set of values for the abundance levels.

The performance of the networks as classifiers of river water quality was not as good as that achieved using Bayesian inference (Walley *et al.* [22]) or Dempster-Shafer reasoning (Boyd *et al.* [2]), but this was only to be expected since these knowledge-based systems incorporated the results of an extensive knowledge acquisition exercise. All three studies used the same set of field data and were therefore comparable in that respect, but the neural networks depended entirely upon the field data for their 'knowledge'. Given a larger set of good quality training data it is confidently anticipated that the network's performance will improve to a level comparable to that of the knowledge-based systems. In fact, the performance of the BMWP score and ASPT on this same set of data (Walley [21]) was similar to that of the networks. Thus the networks have already achieved a level of performance similar to that of nationally recognised systems in the UK.

When trained on relatively small training sets, as in this study, neural networks suffer from over parameterisation and data over-fitting becomes far more likely (Moody [16], Baum and Haussler [1]). Thus they lose much of their inherent ability to generalise and are influenced far more by outliers in the training set. One advantage of the Bayesian and Dempster-Shafer models is that they are capable of identifying conflicting evidence in the field data (e.g. mis-identified taxa) and offer the possibility of correcting for this prior to classifying the sample. In the case of neural networks this can presently only be achieved by pre-processing the data.

However, a current limitation of the knowledge-based systems is that they rely on the samples being taken from riffle biotopes (i.e. fast flowing sections with an eroding sub-stratum). It is within these areas that the water quality is the major factor affecting community composition, and hence these sites are the most reliable for the direct interpretation of the water quality. In fact, one problem relating to the data set used throughout these studies has been that the authors have not been certain as to what extent the samples were taken from riffle biotopes, since they were derived from historic records. Consequently, this may have contributed to the error in some of the classifications. Extension of the knowledge-based systems to overcome this limitation would not necessarily improve the classification of water quality, because sampling from mixed biotopes introduces many more extraneous factors. However, interest in the development of systems to classify environmental quality is growing, and these must, by their very nature, be capable of interpreting composite samples from a wide range of biotopes. Given an appropriate training set, with catchment, climatic and site-specific geomorphic data in addition to mixed-biotope biological data, a neural network could classify environmental quality, and possibly water quality, taking into account the various physical and other factors influencing the biology.

Despite the limitations of the data, this study has demonstrated that neural networks offer a powerful new approach to the classification of river water quality based on the direct interpretation of biological data. Consequently, the authors are currently compiling a much larger and more reliable training set on which to develop their future work. This work will centre around the encoding of available *a priori* information into the network, and the use of neural models with fewer parameters. Geomorphologic and seasonal factors will be included as input to allow better representation of regional and seasonal variations in community structures.

CONCLUSION

A simple neural network model has been applied to the problem of directly interpreting benthic macro-invertebrate data into river water quality class. It has been demonstrated that it is possible to produce a mapping from the sample data to a biologically-based water quality classification system. This was achieved despite the inadequacy of the training data and the subsequent limitation which this placed on the complexity of the network used. The study has thus demonstrated that neural networks have considerable potential for use as classifiers of river water quality from biological data..

ACKNOWLEDGEMENT

The work described in this paper was carried out, in part, under a UK Science and Engineering Research Council studentship.

REFERENCES

1. Baum, E. B. and Haussler, D. 'What Size Net Gives Valid Generalisation ?' *Neural Computation*. Vol. 1, pp. 151-160, 1989.

2. Boyd, M., Walley, W. J. and Hawkes, H. A. 'Dempster-Shafer reasoning for the biological surveillance of river water quality.' *Proc of the 2nd Int. Conf. on Water Pollution*. Milan, June 1993.

3. Chandler, J.F. 'A biological approach to water quality management' *Water Pollution Control*. Vol 69(4), pp. 415-422, 1970.

4. Department of Environment. '*Assessment and Presentation of the Biological Qualities of Rivers in Great Britain*.' Final report of the Biological Monitoring Working Party of the Standing Technical Advisory Committee on Water Quality, 1980.

5. De Pauw, N. and Hawkes, H. A. 'Biological Monitoring of River Water Quality.' *Proc. of the Freshwater Europe Symposium.* Aston Univ, Birmingham, Feb. 1993.

6. Hawkes, H.A. 'An ecological basis for the biological assessment of river pollution' *Chemistry and Industry*, pp. 437-442, 1964.

7. Hawkes, H. A. 'Water Quality: Biological Considerations.' *Chemistry and Industry*, pp. 437-422, 1974.

8. Hawkes, H.A. 'Invertebrates as indicators of river quality' Chapter 2, *Biological Indicators of Water Quality*, eds. James, A. and Evison, L., John Wiley and Sons, Chichester, New York, Brisbane and Toronto, 1979.

9. Hawkes, H.A. 'Biological surveillance of rivers' *Water Pollution Control* Vol 81(3), pp. 329-342, 1982.

10. Hawkes, H.A. and Hughes, J.G. (Eds) 'Biological surveillance of river water quality' Section K, jointly with Section D, *Trans. Proceedings of the British Association for the Advancement of Science*, Aston University, Birmingham, 1977.

11. Hellawell, J. M. *Biological surveillance of rivers.* Water Research Centre: Stevenage, England, 1978.

12. Hertz J. Krogh, A. and Palmer, R. *Introduction to the Theory of Neural Computation.* Addison-Westley, 1991.

13. Kolkwitz, R. and Marsson, M. 'Oekolgie der pflanzlichen saprobien' *Ber. dtsch. bot. Ges.* Vol. 26A, pp505-519, 1908.

14. Kolkwitz, R. and Marsson, M. 'Oekologie der kerischen saprobien' *Rev. Ges. Hydrobiol. Hydrogr.* Vol. 2, pp126-152, 1909.

15. Metcalfe, J. L. 'Biological Water Quality Assessment of Running Waters Based on Macroinvertebrates Communities - History and Present' *Environmental Pollution.* Vol. 60, pp101-139.

16. Moody, J. E. 'The Effective Number of Parameters: An Analysis of Generalization and Regularization in Nonlinear Learning Systems' *Advances in Neural Information Processing Systems 4*, 1992.

17. Moss, D., Furse, M. T., Wright, J. F. and Armitage, P. D. 'The Prediction of the Macroinvertebrate Fauna of Unpolluted Running Water Sites in Great Britain using Environmental Data' *Freshwater Biology.* Vol. 17, pp41-52, 1987.

18. National Rivers Authority. *Proposals for Statutory Water Quality Objectives.* Water Quality series No. 5. National Rivers Authority: ISBN 1-873160-10-0, 1991.

19. Rumelhart, D. E., Hinton, G. E. and Williams, R. J. 'Learning Internal Representations by Error Propagation' *Parallel Distributed Processing,* Vol 1, pp. 318-362, 1986.

20. Sladecek, V. 'Continental systems for the assessment of river water quality' *Biological Indicators of Water Quality*, eds. James, A. and Evison, L., John Wiley and Sons, Chichester, New York, Brisbane and Toronto, 1978.

21. Walley, W. J. 'Artificial intelligence in river water quality monitoring and control' *Proc. of the Freshwater Europe Symposium*. Aston Univ., Birmingham, Feb 1993.

22. Walley, W. J., Hawkes, H. A. and Boyd, M. 'Application of Bayesian inference to river water quality surveillance' *Proc. of the 7th Int. Conf. on Artificial Intelligence in Engineering*, Waterloo, Canada, July 1992.

23. Woodiwiss, F. S. 'The Biological System of Stream Classification Used by the Trent River Board' *Chemistry and Industry*. Vol. 11, pp. 443-447, 1964

24. Wright, J. F., Armitage, P. D., Furse, M. T. and Moss, D. 'A New Approach to the Biological Surveillance of River Quality using Macroinvertebrates' *Verhandlungen Internationale Vereinigung Fur Theoretiche une Angewandtle Limnologie*. Vol. 23, pp. 1548-1552, 1988.