Pergamon

PII: S0043-1354(98)00333-9

Wat. Res. Vol. 32, No. 11, pp. 3500–3501, 1998

Printed in Great Britain
0043-1354/98/$19.00+0.00

COMMENT

Comment on "Neural network predictors of average score per taxon and number of families at unpolluted river sites in Great Britain", by W. J. Walley and V. N. Fontama *Wat. Res.* **32**(3), 613–622

This study is a valuable examination of the difficulties posed by efforts to assess the biological conditions of rivers. The authors have presented a wide range of biological and computational information as the background for their efforts to predict the Biological Monitoring Working Party (BMWP) measures of biological conditions, which are based on observations of macroscopic invertebrates (such as worms and snails) that live in the streams and rivers of England and Wales. As the chief effort of this paper is to compare the results of the authors' artificial neural network models to the results from a previously developed model called RIVPACS III, using data from the Institute of Freshwater Ecology, and since neural networks are not widely used at present in environmental engineering research, a few comments about neural network modeling seem appropriate.

Artificial neural networks, especially those "trained" (i.e., fitted to data) by the backpropagation algorithm, have become popular in many areas of computation and data analysis recently because, being nonlinear, they provide a wider range of model functions than linear regression does, and the training procedures can often converge to good fits when the users have little knowledge of details of the data. Although they were inspired by analogies to biological researchers' concepts of the behavior of living nervous systems, they are formulas that can be mathematically studied independently of the original motivation, and there is no need for an ecologist or engineer to be concerned about whether backpropagation is a realistic model of learning in humans or any other type of organism.

In particular, most implementations of backpropagation rely on having $f(x)=1/(1+\exp(-x))$ as the transfer function of the processing elements (PEs), because the derivative of $f(x)$ is $f(x)\times(1-f(x))$, and this makes it easy to program the use of the chain rule in the algorithm, but this function is not necessarily a highly precise approximation to the actual nonlinear response of a living neuron to differing intensities of stimulation. Likewise, real nerve cells respond to a combination of the inputs that they receive on many dendrites, but the actual electrochemisty of the stimulation process may depart from the simple weighted summation used in the mathematical model. Perhaps some of this should have received a little more emphasis.

Since fitting data to neural networks is outside the conceptual framework that supports conventional procedures for least-squares regression, some aspects of the process are presently done by guesswork and experience, or are chosen by empirical testing methods. The authors did empirical testing to determine the most important environmental variables for predicting NFAM and ASPT, but they do not mention any tests on alternatives for other parts of their network structure. Did they try using only one hidden layer instead of two, or any other number of nodes in either hidden layer instead of six?

The number of nodes per layer and the number of layers has a strong effect on the statistical quality of the resulting fit, since the number of weights in the model is the sum of the products of the numbers of nodes in consecutive layers. Thus, for the networks with 13 inputs, the number of weights is $13\times6+6\times6+6\times1=120$, and for the 5-input and 7-input networks selected as the best predictors of ASPT and NFAM, respectively, the numbers of weights are 72 and 84, respectively. As the training sets F1 and F2 were each half of the 614-record database of sites in Great Britain that are believed to be little polluted, each fit of one of the models was to a set of 307 records. Goodman (1997) recommends (in a text that is not only a user's manual for backpropagation software developed at the University of Nevada at Reno but a discussion of the goals and methods of statistical analysis in general) that the number of data values should be at least ten times the number of weights, to avoid the problem of overfitting. This reasoning suggests that using only one hidden layer of six nodes would have made the number of weights more compatible with the available data. It also suggests that if the modular network with three experts had "far more" parameters than the chosen backpropagation network then the overfitting problem could have been much worse, or if there were more parameters than data points then in principle there could have been nonunique perfect fits to the data.

Another point where current neural network practice appears cruder than the practice of least-squares fitting is the assessment of the significance of variables. The F test has long been established as the way to

determine whether including a variable makes a significant difference in the quality of a fit, assuming that the data were sampled from a population with a Gaussian distribution, and taking into account the difference between the number of parameters in the model and the number of data values, usually called the number of degrees of freedom. Evidently, the F test does not apply to tests of removal of parameters from neural network models, but confidence levels and tests of significance have also been developed for correlation coefficients, based on numbers of degrees of freedom, so if one wants to determine whether removing parameters makes a significant difference in the quality of the fits, as in Table 2, it might be better to consider significance tests instead of simply calculating percentage changes in correlation coefficients. What would be the authors' response?

These computational considerations do not address problems cited by the authors in their examination of the shaky conceptual basis for the data on which they worked. They note that variations in the sampling effort at different sites are likely to have contributed to the fact that neither RIVPACS III nor the neural networks were able to predict NFAM as well as they predicted ASPT, and that all of the sites are subject to natural temporal variations that makes NFAM and ASPT stochastic variables (Iranpour *et al.*, 1988), instead of the constants assumed by the prediction process. Moreover, the present list of environmental variables omits geological information that is known to be important, and they believe that it could also be useful to know the amount of woody litter at a site.

However, improving the input data must be considered to be complementary to the task of improving computational methods, since poor computation would largely if not entirely negate the value of better data. In particular, additional mathematical thought may be needed to deal with random aspects of the measurements. Would the authors agree that considering established approaches to confidence limits would be a useful starting point?

The authors have made much effort to get the most of a combination of imperfect data and an analysis method that is still new. We believe that there is a great future for improved methods of data analysis in environmental research, and we hope that the authors will continue their work in this direction.

REFERENCES

Goodman P. (1997) NevProp3 User Manual. Available from the University of Reno at ftp://ftp.scs.unr.edu/pub/cbmr/nevpropdir.

Hykin S. (1994) Neural Networks: A Comprehensive Foundation. Macmillan.

Iranpour R., Chacon P. and Kac M. (1988) Basic Stochastic Processes. Macmillan Publishing Co., New York, N.Y.

R. IRANPOUR[1]Ⓜ
W. PROSKUROWSKI[2]
D. MILLER[3]
M. SAADATPOUR[4]

[1] *Wastewater Engineering/Applied Mathematics*
PO Box 806
Culver City, CA 90232
USA

[2] *Applied Mathematics/Computer Science*
University of Southern California
Los Angeles, CA
USA

[3] *Applied Mathematics*
Tech RCT Inc.
Los Angeles, CA
USA

[4] *Civil and Environmental Engineering*
Isfahan University of Technology
Isfahan
Iran

Pergamon

PII: S0043-1354(98)00334-0

Wat. Res. Vol. 32, No. 11, pp. 3502–3503, 1998

Printed in Great Britain
0043-1354/98/$19.00 + 0.00

AUTHORS' REPLY

The precise relationship between the behaviour of artificial neural networks and that of real neural networks was never thought to be relevant to the paper. The authors' only reason for using the brain analogy when explaining the structure and function of artificial neural networks was to help the readership, primarily biologists in this case, to gain an adequate understanding the analytical technique being used. Feedback to date has indicated that this approach was found to be helpful. We were not arguing that backpropagation is a realistic model of learning in humans. Indeed, we agree that although inspired by the behaviour of living nervous systems, artificial neural networks are at best very simple models of them, and that many modern networks have strayed far from the brain based concept. The backpropagation algorithm, for instance, is a semi-parametric method that bears more resemblance to statistics than biology. It uses gradient descent algorithms, which are widely known in statistics, to minimise the error function. The similarity between these algorithms and the processes of biological neurons is indeed scant.

Before commenting on the remarks made about the topology of the networks used and the adequacy of our data set, it may be helpful to give a little more background to the study. It was carried out as a small part of an investigation into possible applications of neural networks in river quality surveys. The main body of the work used a database consisting of 12,076 biological samples from 6,038 sites covering all river qualities. Several different networks were trained using this large database, but these were not reported on in the paper because they represent an approach to data interpretation that takes a more holistic view of "clean" and "dirty" water biology (Walley and Fontama, in press; Walley *et al.*, 1998) than the one in question. The prediction of the average score per taxon (ASPT) and number of families (NFAM) at unpolluted (or unstressed) sites is a key component of the 'reference state' approach to the classification of river quality. The authors are not advocates of this approach, but advocates of the more holistic approach. Nevertheless, the 'reference state' approach, based on RIVPACS, is presently the industry standard in the United Kingdom. As explained in the paper, RIVPACS is a statistical procedure involving three consecutive mathematical mappings: 13 environmental characteristics to one of 35 site types; site type to predicted unstressed biological community; and biological community to predicted ASPT and NFAM.

The purpose of our study was to carry out a like-for-like comparison between RIVPACS predictions and those of neural networks designed to provide a single non-linear mapping from the 13 environmental characteristics to the predicted ASPT and NFAM. To ensure the like-for-like comparison we had to use the same database as that used to develop RIVPACS. Yes, we do have doubts about its adequacy for the purpose, but these are equally relevant to the development of RIVPACS as to the neural networks. The number of estimated parameters used by RIVPACS has not to our knowledge ever been stated, but we would be very surprised if it were less than that used by the networks.

The final choice of the topology of the networks was made after studying the performance of many different topologies. We did not feel that we could contribute anything new by detailing the performance of different architectures since this aspect of neural networks has been extensively reported in the literature (Bishop and James, 1993; Bishop, 1995).

We were very conscious of the danger of over-fitting the data, as we made clear in the paper, which was why we used cross-validation to prevent over-training and to provide data for independent performance tests. The success of this approach is reflected in the fact that the modular networks, with their larger number of parameters, did not out-perform the much simpler backpropagation networks that were finally adopted. In addition, the fact that there was very little difference between the correlations achieved by the independent (F1/F2 and F2/F1) and the dependent (F1/F1 and F2/F2) ASPT models showed that there was little danger of overfitting in this case. In the case of the NFAM models the difference was much greater, thus indicating that there would have been an over-fitting problem had we not used cross-validation. The extent to which this problem affected the RIVPACS model, which was not cross-validated, we do not know. We are not convinced that statistical practices are necessarily any less crude than current neural network practices. In our experience many statistical models are not cross-validated and some are based on shaky assumptions about distributions and independence. Clearly, there is a need to promote best practices in both fields.

It is true that over-fitting can be avoided by using a large number of records, but, as explained earlier, this was not an option in our case. Regularisation offers another means of avoiding the over-fitting of

neural networks. This is done by adding a term to the error function which reduces the complexity of the model and results in a smoother mapping that fits less of the noise in the training data.

As regards the removal of irrelevant variables, we feel that impact tests, based on changes in the correlation coefficient, provide a simple and effective means of ranking variables in terms of their importance. However, we do agree that appropriate significance tests can offer a more systematic means of achieving this, but the F-test and t-test can be a bit limiting as they assume that the data are drawn from a normal distribution. There are other good techniques that can be used to select the important variables. These include the branch and bound method, the sequential forward selection and sequential backward selection techniques (Bishop, 1995).

We agree that it is desirable to attach confidence limits to predictions, whether they be produced by neural networks or statistical methods. This has been the subject of much research within the neural network community (Neal, 1994; Mackay, 1992). Bayesian methods have been applied to neural networks with the result that error bars can now be assigned to each prediction. Thus, we accept that the adoption of established statistical methods, where applicable, is definitely a step in the right direction.

We agree with the comment about the shaky conceptual basis of the data. Our investigation of the sources of error and bias was designed to examine this issue. Indeed, we would go further by saying that the whole basis of the 'reference state' approach is rather shaky. Firstly, there is the difficulty of defining and then finding sites that can be classified as reference sites (i.e. unpolluted or unstressed). Some countries, like The Netherlands, would find it exceedingly difficult to identify any such reference sites. Secondly, the relationship between the composition of an aquatic community and the river's state of stress is highly non-linear, and the notion that the state of stress of a river can be properly defined by simple linear metrics relating the actual community to the reference community is in our view naive. In addition, we concluded that predictions of NFAM were too unreliable to be used in the formulation of such metrics. Despite the potential improvements to the method that have been identified, we are of the view that the problems facing the 'reference state' approach may well prove so insurmountable as to require a re-think along the more holistic lines mentioned earlier. However, these wider issues were not discussed in the paper because we took the view they were beyond its intended scope.

Bishop C. M. (1995) *Neural Networks for Pattern Recognition*. Oxford University Press.

Bishop C. M. and James G. D. (1993) Analysis of multiphase flows using dual-energy gamma densitometry and neural networks. Nuclear Instruments and Methods in Physics Research A327, 580–593.

Mackay J. D. C. (1992) A practical Bayesian framework for backpropagation networks. *Neural Computation*. **4**(3), 448–472.

Neal R. M. (1994) Bayesian learning for neural networks. PhD thesis, University of Toronto, Canada

Walley W. J., Fontama V. N. and Martin R. (1998) Applications of artificial intelligence in river quality surveys. Environment Agency, R&D Technical Report E52. Foundation for Water Research (In press).

Walley W. J. and Fontama V. N. (in press) New approaches to the biological classification of river quality based upon artificial intelligence. In *RIVPACS International Workshop, Oxford, 16–18 September 1997*. Eds. Furse M. T. and Wright J. F., Freshwater Biological Association, Ambleside. (In press).

W. J. WALLEY✉
V. N. FONTAMA

Staffordshire University
School of Computing
The Octagon
Beaconside
Staffs ST18 0DG
UK