The extrapolation of artificial neural networks for the modelling of rainfall–runoff relationships

P. Hettiarachchi, M. J. Hall and A. W. Minns

ABSTRACT

The last decade has seen increasing interest in the application of Artificial Neural Networks (ANNs) for the modelling of the relationship between rainfall and streamflow. Since multi-layer, feed-forward ANNs have the property of being universal approximators, they are able to capture the essence of most input–output relationships, provided that an underlying deterministic relationship exists. Unfortunately, owing to the standardisation of inputs and outputs that is required to run ANNs, a problem arises in extrapolation: if the training data set does not contain the maximum possible output value, an unmodified network will be unable to synthesise this peak value. The occurrence of high magnitude, low frequency events within short periods of record is largely fortuitous. Therefore, the confidence in the neural network model can be greatly enhanced if some methodology can be found for incorporating domain knowledge about such events into the calibration and verification procedure in addition to the available measured data sets. One possible form of additional domain knowledge is the Estimated Maximum Flood (EMF), a notional event with a small but non-negligible probability of exceedence. This study investigates the suitability of including an EMF estimate in the training set of a rainfall–runoff ANN in order to improve the extrapolation characteristics of the network. A study has been carried out in which EMFs have been included, along with recorded flood events, in the training of ANN models for six catchments in the south west of England. The results demonstrate that, with prior transformation of the runoff data to logarithms of flows, the inclusion of domain knowledge in the form of such extreme synthetic events improves the generalisation capabilities of the ANN model and does not disrupt the training process. Where guidelines are available for EMF estimation, the application of this approach is recommended as an alternative means of overcoming the inherent extrapolation problems of multi-layer, feed-forward ANNs.

Key words | artificial neural networks, rainfall-runoff modelling, extrapolation, domain knowledge

P. Hettiarachchi
Office of the Deputy Director of Irrigation,
Kaluwella, Galle,
Sri Lanka

M. J. Hall
UNESCO-IHE Institute for Water Education,
POB 3015, 2601 DA, Delft,
The Netherlands

A. W. Minns (corresponding author)
WL | Delft Hydraulics,
POB 177, 2600 MH, Delft,
The Netherlands
Tel: +31-15-2858952
Fax: +31-15-2858582
E-mail: *tony.minns@wldelft.nl*

INTRODUCTION

Over the last decade, there has been a marked increase in the number of applications of Artificial Neural Networks (ANNs) for the solution of hydrological problems. Reviews by ASCE (2000*a*, *b*) and Dawson & Wilby (2001) have demonstrated the wide variety of modelling problems for which ANNs might be used to advantage. To date, the majority of studies have been directed towards the relationship between precipitation and streamflows.

A significant problem with ANN rainfall–runoff models is their inability to extrapolate. The input to each node in the hidden layer of a multi-layer perceptron consists of a summation of the products of the weights assigned to each connection from the input layer and the inputs themselves. This sum is then passed through an activation function, generally of sigmoidal form, such that the outputs from any layer are constrained to a range (say) from zero to one.

The result is that the extremes in the training data set are scaled to the zero–one interval. If then (say) the largest values in the testing data set exceed those in the training set, then the former will not be reproduced by the model. Some latitude is provided by varying the range of standardisation to, for example, 0.2 to 0.8, but the ability of the ANN to extrapolate remains limited (see Minns (1996) and Minns & Hall (1996)). Although the use of a linear activation function in the output layer may appear to offer a degree of extrapolation beyond fixed bounds, in practice the amount of extrapolation is then limited by the saturation of the nodes in the hidden layer. More recently, a more elaborate method of scaling the activation function to leave "room" for extrapolation has been proposed by Varoonchotikul *et al.* (2002) and Varoonchotikul (2003), but at the expense of introducing an additional, initially arbitrary, parameter. In addition, Imrie *et al.* (2000) have proposed that, in order to assist the ANN to extrapolate, a *guidance system* should be added to the output layer. However, the procedure suggested for determining an appropriate form for the guidance system involves the use of the testing data set, thereby disrupting the cycle of training and independent verification and introducing even greater dependence on the available input and output records.

Bearing in mind that ANNs are knowledge encapsulators, an alternative approach to extrapolating an ANN-based rainfall–runoff model might involve the introduction of extra domain knowledge about the catchment of interest over and above that contained in the available data sets. For example, in determining the capacities of spillways for large dams whose failure may involve heavy loss of life, design engineers have adopted as their standard an Estimated Maximum Flood (EMF), which is intended to approximate the physical upper limit of catchment response to the Probable Maximum Precipitation (PMP). More realistically, the EMF might be defined as the flood with a very small, but non-negligible, probability of annual exceedence. The importance of EMF estimation to public safety, and the position of the design engineer in legal terms, has resulted in many countries adopting standard procedures for EMF estimation. In the United Kingdom, for example, the EMF is estimated by a procedure that is an extension of the unit hydrograph plus design storm approach in the UK Flood Studies Report (NERC 1975), which has been perpetuated in the UK Flood Estimation Manual (Houghton-Carr 1999). Although based upon the well-tried unit hydrograph and loss function methodologies, the computation of the EMF is founded upon an extensive regionalisation exercise of data from UK catchments. Similarly, the associated PMP storm profile is based upon the extrapolation of UK rainfall statistics. The results may therefore be regarded as complementary to any catchment-specific rainfall and flow data, thereby providing a ready source of additional domain knowledge. Such knowledge is easily incorporated into an ANN modelling exercise by adding an EMF hydrograph, or that of any appropriate long return period design event, together with the associated storm profile, into the training data set. Since the EMF or any long return period event may represent a peak flow well in excess of those in the available flow record, the question arises as to what effect its inclusion might have on the training process and the ability of the ANN to generalise. A series of numerical experiments were therefore developed to investigate the efficacy of this approach, and the results are reported in this paper.

HYDROLOGICAL DATA

Hydrological data were obtained for six catchment areas in the south west of England (see Table 1). The data for individual storm events were taken from the Representative Basin catalogue for the United Kingdom, an archive of flood events maintained by the Centre for Ecology and Hydrology (formerly the Institute of Hydrology), Wallingford, UK.

The EMF hydrograph for each of the six catchments was constructed using the procedure summarised in the Flood Studies Report (NERC 1975, ch. 6). In brief, a PMP profile is constructed using mapped values of the 2-h and 24-h estimated maximum rainfall. The storm duration, D, is a function of the (mapped) average annual rainfall and the time to peak of the 1-h unit hydrograph, T_p. A series of tabulated growth factors and areal reduction factors are used to construct a PMP depth–duration curve from which the D-h total is interpolated. An additional allowance is made for snowmelt runoff at a rate of 42 mm/d. The estimated maximum D-h rainfall total is distributed symmetrically such that the estimated maximum total is contained within each sub-duration centred on the peak.

Table 1 | Summary of the catchment areas and pertinent flow characteristics. Mean annual floods (MAF) from Rees *et al.* (1993); estimated maximum floods (EMF) by calculation

Catchment	Area (km^2)	Years of data	Number of events	MAF peak (m^3/s)	EMF peak (m^3/s)
Dart	248	1963–1969	22	228	1867
Thrushel	112.7	1971–1985	37	54	760
Plym	79.2	1971–1975	15	–	665
Yealm	54.9	1965–1976	17	21	446
East Dart	21.5	1964–1976	22	44	335
Swincombe	14.2	1963–1968	18	–	233

A design percentage runoff for the EMF is computed from the sum of three elements. Firstly, an allowance is made for "standard" conditions, dependent on a (mapped) soil index and a proportion of urban area. Antecedent conditions are then allowed for using a catchment wetness index, assuming that the *D*-h PMP occurring in the middle of a 5*D*-h storm. Finally, a contribution dependent on storm size is added that increases linearly with the total depth of precipitation. The hyetograph of effective rainfall is then convolved with a triangular unit hydrograph whose proportions are entirely dependent on the estimated time to peak. An allowance for baseflow, based on the catchment wetness index and the net 1-d, 5-year rainfall total, is then added to the direct runoff hydrograph to produce the EMF hydrograph. This synthetic event is then inserted into the time series of existing rainfall–streamflow data for the selected catchments. The peaks of the EMFs for each of the six catchments are summarised in Table 1.

NEURAL NETWORK MODELLING

In all cases, a multi-layer perceptron (MLP) ANN was employed for rainfall–runoff modelling, with the weights determined by error back-propagation. Sigmoidal activation functions were used at all nodes in the hidden and output layers. Before presentation to the networks, all series of events were converted into time series. In some cases, the serial order of the recorded events was changed to obtain closer coincidence between the end of one recorded event and the start of the rising limb of its successor. Where the flow discrepancies between recessions and rising limbs were too large, the recorded recessions were extended artificially, assuming the logarithms of flows were reducing linearly. The sections of the rainfall time series corresponding to the flow recessions were infilled with zeros.

The configurations of the ANN models were determined initially by trial and error, with different selections of inputs and different numbers of hidden nodes. The best results in this study were obtained with five inputs: two antecedent rainfalls, the concurrent rainfall and two antecedent flows. In all cases, the output consisted of one value of concurrent flow. For these trials and for all subsequent experiments, the available data were divided into three subsets. The first subset, comprising some 40% of the data, was used for training, with another 20% reserved for cross-validation and the remaining 40% for testing. The networks are initially run with the training subset, the back-propagation algorithm ensuring that the mean square error between observed and computed outputs decreases with repeated presentations ("epochs") of the input data. However, every 100 epochs, the mean square error of the cross-validation subset is checked. Once the latter measure reaches a minimum, the training is terminated, thereby ensuring that the ANN does not "overlearn" and retains an ability to generalise beyond the events in the training data subset. This facility was a standard option in the Neurosolutions software that was employed in this study.

There were three phases to the experiments that were undertaken with the data sets from all six drainage areas. For convenience in presentation, only the results from the Thrushel catchment will be presented in detail, since comparable outcomes were obtained with the other five. In addition, although the need to address more than one index of goodness-of-fit is acknowledged, for brevity the results presented are confined to graphical comparisons and coefficients of correlation between the observed and computed flow hydrographs. The three phases of experimentation may be summarised as follows:

Phase 1: A confirmation of the ability of ANNs to identify usable relationships between the flows and the selected input variables, during which the highest (peak) flow on record was included in the training data subset.

Phase 2: An illustration of the problems of extrapolation and the impaired ability of the ANN to generalise that accrue from the highest (peak) flow on record being included in the testing data subset. For this purpose, the 40% of the total available data that had been used for testing in Phase 1 became the training data subset, with the cross-validation data subset remaining unchanged.

Phase 3: The inclusion of the EMFs for each catchment in the training data subsets in order to evaluate any problems which may arise in training in the presence of such extreme events, and to assess the ability of the trained network to generalise.

MODELLING RESULTS

The Phase 1 training and testing procedure produced similar results for each catchment. For example, for the River Thrushel, the largest recorded peak flow of 123.7 m^3/s was contained in the training subset and the largest peak flow in the testing subset was 61.5 m^3/s. The trained ANN had no problems in learning to reproduce the catchment behaviour with high accuracy. The overall reproduction of the hydrographs was good, with the coefficients of correlation between the computed and observed data being 0.994 for the training and 0.988 for the testing subsets. The maximum absolute errors for both training and testing of the Thrushel data were both in the region of 11–12 m^3/s, although such figures, being derived from the concurrent ordinates, reflect small timing errors as well as errors in magnitude.

When the training and testing subsets were interchanged in Phase 2, the reduction in the coefficients of correlation was marginal for all catchments, but as could be anticipated, the ANN models now failed to reproduce the maximum values in the testing subsets. For example, for the River Thrushel, the threshold of about 65 m^3/s imposed by the training data meant that it was intrinsically impossible for the trained ANN to produce an output value greater than 65 m^3/s, no matter how large the rainfall inputs were as shown in Figure 1(a). These results demonstrate once more the exceptional performance of ANN models to reproduce the rainfall–runoff relationship for a given catchment with well-posed input and output data sets and highlight the poor extrapolation properties of these models for the unwary user.

The objective of the study was then to improve the extrapolation performance of the ANN without decreasing the level of performance already demonstrated in Phase 1. In Phase 3, the estimated maximum rainfall profiles and EMF hydrographs derived from application of the Flood Studies Report procedure were added to the original training subsets of all six catchments. The EMF peaks (see Table 1) represent events that vary between 6 and 20 times the magnitude of the largest peaks contained within the period of available record, and therefore present a considerable challenge in training an ANN model. Taking again the example of the River Thrushel, the learning process was slowed considerably, although after some 10,000 epochs the EMF peak had been successfully reproduced. The training continued to improve, but even after 60,000 epochs, the performance in testing was notably poorer than before the addition of the EMF. In particular, the hydrograph recessions were consistently overestimated in the testing data set, even though reproduction of high flows was good. The coefficients of correlation between observed and computed flows for the Thrushel were 0.992 for the training data set but only 0.909 for the testing data set. These tests were repeated with the data from the Dart and East Dart with very similar results.

Clearly, the ANNs are capable of capturing the high flow behaviour of the catchments when the EMFs are introduced, but training tends to be dominated by the EMF at the expense of reduced attention to low-flow behaviour. This obviously does not satisfy the objective stated above. A method is therefore required to reduce the domination of the EMF in the training procedure whilst maintaining the additional information contained within. One possible

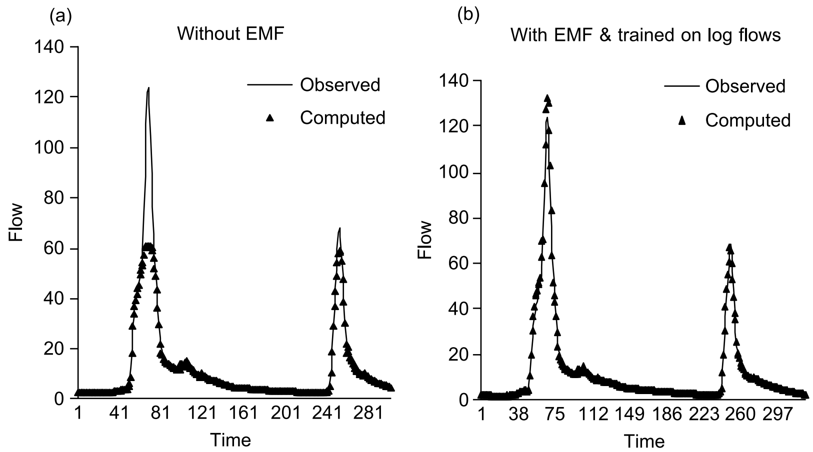

Figure 1 | Comparison between two storm events for the River Thrushel (a) trained with the maximum recorded peak flow rate in the testing data subset and (b) trained with the estimated maximum flood included in the training data set.

means of improving the training is to reduce artificially the range of events by transformation to the logarithms of flows. The ANNs are then trained with rainfall inputs as observed but with log-transformed flows.

For all of the catchments, the benefit of the transformation in improving the reproduction of the observed (log) flows was evident. For the River Thrushel, the correlation coefficients between observed and computed (log) flows were 0.996 in training and 0.988 in testing, easily comparable with those obtained in the Phase 1 tests. The training times for the ANNs were also considerably less than using the non-transformed inputs. Indeed, over all six catchments, correlation coefficients in either training or testing never fell below 0.98. In addition, comparisons were made between the observed flows and the back-transformed model outputs. The back-transformation has the effect of exaggerating the discrepancy in the training of the EMF, resulting in a correlation coefficient for the (non-transformed) training subset of 0.961, although that for the testing subset was higher at 0.992.

The point is further emphasised in Figure 1, which shows two individual hydrographs from the testing data subset for the River Thrushel from the Phase 2 and Phase 3 experiments. With the peak flow in the testing data subset and no EMF in the training subset, the ANN is unable to reproduce flows above about 60 m^3/s. However, with the EMF included and the training performed on the logarithms of flows, the larger of the two events is successfully captured, if not slightly overestimated.

The results for the Plym, East Dart and Swincombe rivers showed a similar reduction in the correlation coefficient for the training subsets. In contrast, for the Yealm, both correlation coefficients were 0.989, and for the Dart, the coefficient for the testing subset fell to 0.954 compared with 0.977 for the training subset. In all cases, the maximum absolute errors were associated with the EMF peak ordinates, but in percentage terms the discrepancies were small, amounting to some 4% in the case of the Dart. Moreover, for all six catchments, the low flow recessions were extremely well reproduced.

Thus a method has been introduced in which the extrapolation properties of a "standard" MLP-type ANN have been significantly improved for rainfall–runoff modelling by adding hydrological domain knowledge to the training data set without compromising either the measured data values themselves or the architecture of the ANN.

CONCLUDING REMARKS

The three sets of numerical experiments that were carried out on over 130 individual flood events for six different catchments in the south west of England have demonstrated once again that ANN-based rainfall–runoff models are capable of reproducing recorded behaviour to a high degree of fidelity. The Phase 2 results also provide a further illustration (if one were required) of the dangers of extrapolation of ANNs beyond the range of the training data set. The benefits of incorporating additional domain knowledge on catchment response into the training process by way of estimated maximum rainfall profiles and EMF flood hydrographs are clearly evident from the Phase 3 results. With the additional step of applying a logarithmic transform to the flow data prior to training, the fit of the computed to the observed flows was comparable to that obtained under the artificially well-posed conditions of the Phase 1 experiments. Moreover, when the ANN that was trained with the EMF on log flows was tested on the training data without the EMF, the coefficients of correlation between observed and computed flows in both training and testing could only be distinguished in the fourth decimal place. These results demonstrate that the training process appears not to be affected adversely by the inclusion of an EMF event, and that the ANN obtained from the training has the ability to extrapolate beyond the recorded maxima in the training data. Furthermore, the EMF estimate can be derived using established hydrological techniques, thus reducing the arbitrariness in the generation of "synthetic" training data.

The adoption of this approach is therefore recommended for those areas in which procedures have been established for the estimation of EMF hydrographs. In this way, the inherent extrapolation problems of a neural network based solely on measured data may be overcome using physical insight into the catchment behaviour rather than a mathematical sleight of hand.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the kind cooperation of the Centre for Ecology and Hydrology, Wallingford, in providing the data from the *Representative Basin Catalogue for Great Britain* that were employed in this study.

REFERENCES

ASCE 2000a Task committee on application of artificial neural networks in hydrology, artificial neural networks in hydrology. I: Preliminary concepts. *J. Hydrol. Engng.* **5**, 115–123.

ASCE 2000b Task committee on application of artificial neural networks in hydrology, artificial neural networks in hydrology. II: Hydrologic applications. *J. Hydrol. Engng.* **5**, 124–137.

Dawson, C. W. & Wilby, R. 2001 Hydrological modelling using artificial neural networks. *Prog. Phys. Geogr.* **25**, 80–108.

Houghton-Carr, H. 1999 *Flood Estimation Handbook*, vol. 4 *Restatement and Application of the Flood Studies Report Rainfall-runoff Method* Institute of Hydrology, Wallingford.

Imrie, C. E., Durucan, S. & Korre, A. 2000 River flow prediction using artificial neural networks: generalisation beyond the calibration range. *J. Hydrol.* **233**, 138–153.

Minns, A. W. 1996 Extended rainfall-runoff modelling using artificial neural networks. In *Hydroinformatics '96. Proc 2nd Int. Conf. On Hydroinformatics*, **vol. 1** (ed. in A. Muller), pp. 207–213. Balkema. Rotterdam.

Minns, A. W. & Hall, M. J. 1996 Artificial neural networks as rainfall-runoff models. *Hydrol. Sci. J.* **41**, 399–417.

NERC 1975 *Flood Studies Report*, vol. I *Hydrological Studies* Natural Environment Research Council, London.

Rees, G., Roald, L. & Dixon, J. 1993 *Flow Regimes from International Experimental and Network Data (FRIEND)*, vol. II *Hydrological Data* Institute of Hydrology, Wallingford.

Varoonchotikul, P. 2003 *Flood Forecasting Using Artificial Neural Networks*. Swets and Zietlinger. Lisse.

Varoonchotikul, P., Hall, M. J. & Minns, A. W. 2002 Extrapolation management for artificial neural network models of rainfall-runoff relationships. In: *Hydroinformatics 2002*, vol. 1 *Model Development and Data Management* (Falconer, R. A., Lin, B., Harris, E. L. & Wilson, C. A. M. E (eds.)). IWA Publishing, London, pp 673–678.