A combination of neural networks and hydrodynamic models for river flow prediction

Nigel G. Wright[1], Mohammad T. Dastorani[1], Peter Goodwin[2] & Charles W. Slaughter[2]

[1]School of Civil Engineering, University of Nottingham, Nottingham NG7 2RD, UK
[2]Eco-Hydraulics Research Group, University of Idaho, Boise, Idaho 83712, USA

ABSTRACT: This research has investigated the application of artificial neural networks (ANN) to improve the accuracy of the results obtained from a hydrodynamic model of river flow. The study area was Reynolds Creek Experimental Watershed in southwest Idaho, which has 239 km^2 drainage area and semi-arid climate conditions. Hydrological processes in this catchment are extremely variable because of the high variation in elevation, and consequently climate condition, within the river catchment. After the calibration of a 1D hydrodynamic flood routing model of the main river reach, a MLP neural network model has been adopted to optimise the outputs of the hydrodynamic modelling procedure. Using ANN in this way, the error produced by the hydrodynamic model was predicted and thereby, the results of the model were improved. In addition, the results of a hydrodynamic model affected by the suspension of flow gauging are improved by neural networks. Combination of these two techniques for this specific application uses the potential of both methods and shows a good performance.

1 Introduction

Flood routing procedures to estimate the changes in a hydrograph as it moves along a river is one of the main methods for producing flood warnings and mitigating damage. It can also play a crucial role in other aspects of water resources planning and management. Generally such routing is carried out using hydrological or hydraulic routing methods. These models use information on cross-sections, bed material etc. to construct a model. Once this has been done the model is calibrated by adjusting parameters such as friction factor to ensure that the model results match the measured data. Care must be taken to ensure that the changes to the parameters are not unrealistic and do not go beyond the physically realistic range. In some cases discrepancies between models and measure data are due to omitting aspects of the physical problem either because the phenomena is too complex or because no data is available. An example of the former is Manning's formula which is used to represent bed shear and turbulence. The latter encompasses situations where cross-section data or hydrological data are not available. Rather than taking a calibration parameter and adjusting it to an unphysical value to represent a phenomena that it was never intended to represent, this paper presents work that use a neural network to predict this error based on learning from errors in past model runs. As such the work is making use of the symbolically expressed physical laws (conservation of mass, Newton's Second Law) where they are applicable and augmenting them with sub-symbolic representations based on past experience.

Hydroinformatics 2002: Proceedings of the Fifth International Conference on Hydroinformatics, Cardiff, UK
 ISBN 1 84339 021 3 (set)

During the last decade new machine learning techniques such as artificial neural networks, genetic programming and fuzzy logic have been widely applied to different aspects of water-related modelling such as rainfall-runoff, flow discharge and stage prediction, regional flood frequency, etc. Despite some similarities, each of these new techniques works with its own procedure. Neural networks operate on the principle of learning from a training set. They must be trained with a set of typical input/output pairs of data called the training set. The final weight vector of a successfully trained neural network represents its knowledge about the problem. In general, it is assumed that the network does not have any a priori knowledge about the problem before it is trained. At the beginning of training the network weights are usually initialised with a set of random values. There are a variety of neural network models and learning procedures.

There have been some discussions of the relative merits of ANN and hydrodynamic approaches to predicting river flows. As outlined above this research uses a combination of the two in a new approach to river modelling where the artificial neural networks (ANN) are used to improve the accuracy of the results obtained from a hydrodynamic model of river flow. A MLP neural network model has been adopted to optimise the outputs of the hydrodynamic modelling procedure by prediction of the error in the hydrodynamic model.

2 Study area

The study area for this research is Reynolds Creeks Experimental Watershed (RCEW), a typical intermountain region of the western United States. It is located in the Owyhee Mountains of south-western Idaho with a 239 km^2 drainage area. The main stream flows from south to north at an elevation exceeding 2200m. Most parts of the catchment have rugged topography except the north central part where there is a broad valley floor. Elevation ranges from 1101 m in the north to a peak of 2241 m in the south.

Local slope and aspect strongly influence the hydrology of the catchment by controlling the incoming solar radiation and snow deposition patterns. The Reynolds Creeks Experimental Watershed represents a typical arid and semiarid climate condition where climate elements vary highly from part to part. The climate is largely controlled by the elevation and local topography. Precipitation varies from about 230 mm at the northern lower elevation to over 1100 mm in the higher regions at the southern and southwestern boundaries where 75% or more of annual precipitation occurs as snowfall (Hanson 2000). According to 35 years (1963-1996) of record, annual precipitation has increased 887 mm, or 4.8 times, over a distance of only 17.6 km between two gauging sites located in two different elevations. Based on dew point temperature during storms, 15-55% of the lower elevation precipitation falls as snow, and 60-90% of the higher elevation precipitation falls as snow (Marks et al; 2000). Water from snowmelt is very important, as it is the main source of soil moisture and stream flow during spring and summer. For the outlet of the catchment average annual water yield has been measured as 0.564 m^3/s or 77 mm. The variation of runoff is considerable from year to year. The largest recorded flow at the outlet gauging site is over 107 m^3/s. This occurred on December 23, 1964, due to a rain-on-snow event with a frozen soil surface. Flow measurement in RCEW started at two sites in 1963, at three other sites in 1964, followed by eight additional sites in subsequent years. However, measurements were discontinued later in five gauging stations. Gauging sites were established in sub-catchments ranging in contributing area from 1.03 to 23866 hectares. Quality controlled and validated hourly stream flow data sets are available for 13 sites for the period 1963-1996 (or a subset of that time for some sites) (Pierson et al. 2000).

3 Modelled river reach

In this study the Reynolds Creek main river reach between Tollgate gauging station upstream and the Outlet weir downstream has been modelled. This reach is a perennial stream with 14.082 kilometres length of stream course (the thalweg length is 17.073 km). The downstream end of the reach is a self-cleaning overflow V-Notch (SCOV) weir draining all 23866 hectares of the catchment, and located in a narrow canyon at low elevation (1108 meters above sea level), and about 12 kilometres south of the confluence of Reynolds Creek and the Snake River. The upstream end of the reach is limited to the Tollgate weir. This is a gauging station with a Drop-Box V-Notch weir located at an elevation of 1410 m above sea level. Drainage area above this weir is 5457 hectares, ranging from 1410 m to 2241 m with precipitation mostly in the form of snow.

4 Hydrodynamic modelling

River floods are normally gradually varied unsteady flows and so a time-dependent simulation is required. Further, most computational models used to predict river floods are one-dimensional. DHI's MIKE11 is a professional engineering software package for the unsteady hydrodynamic simulation of flows, water quality and sediment transport in estuaries, rivers, irrigation systems, channels and other water bodies.

A model was constructed using MIKE11 and was calibrated for the reach between Tollgate and Outlet using flow data for a flood wave occurring in the period of February-April 1982. Flow data were entered to the model from Tollgate, Macks Creek and Salmon Creek (three gauging stations upstream side of the outlet weir). These three gauging sites, which drain three main sub-catchments can be used to model the main river reach. All other gauging sites are located in headwater tributaries of the sub-catchments and do not deliver flow directly to the main river. As mentioned earlier the operation of Macks Creek has been suspended since 1991 because of ownership problems. The second step in this research was to evaluate how gauging site suspension affects the procedure of hydrodynamic modelling and how the technique of artificial neural networks can bridge this gap and improve the results of hydrodynamic model affected by this shortage of measured data. Therefore, two different simulations were carried out: the first one uses flow hydrographs from all three gauging stations as boundary conditions of the model, and the second one uses only two stations where gauging was continued (Tollgate and Salmon Greek). 53 cross sections surveyed between two ends of the reach define the topographical characteristics of the river reach in the model. The estimated flow hydrographs at the outlet for the period of February-April 1982 in both first and second simulations have been compared to the measured values. Figures 1 and 2 show hydrographs for March 11 to April 16 for first simulation (with Mack's Creek data) and second simulation (without Mack's Creek data) respectively. As the figures show the estimated values are not enough close to the measured values especially for the peak flow and particularly for the results of second simulation. As a result the output of the model at this stage may not be accurate enough for practical applications.

5 Prediction of error by ANN

To improve the outputs of the hydrodynamic model, an artificial neural network was employed to estimate the errors of modelling results in both the first and second simulations. The artificial neural network architecture used was a three-layer perceptron feedforward neural network. This type of the neural network is usually trained with the backpropagation algorithm. The backpropagation rule propagates the errors through the network and allows adoption of hidden processing elements. The final weight vector of the successfully trained network, which represents its knowledge about the problem, is used to apply to a new set of data to evaluate the performance of the model.

The errors of the hydrodynamic model were calculated using the following formula:

$$E_p = X_{obs} - X_{est}$$

where E_p is the error of estimated pattern, X_{obs} is the observed value and X_{est} is the estimated value. The ANN model was trained using Tollgate measured flow and MIKE11 output for outlet as input to the ANN model and error of MIKE11 estimation (E_p) as the output of the neural network model.

Hourly flow data for the period of February 1 to April 30, 1982 were used for this simulation. Prediction of the error for the results of the first and second hydrodynamic simulations was carried out separately. The input to the network was flow data from the Tollgate station and the estimations from MIKE11 for the outlet. The output from the network was the error of MIKE11 estimations for outlet (E_P). Data was divided into three parts for training, testing and cross validation purposes. February 1 to March 10 was used to train the model to establish the relationship between input and output patterns, while the data of March 11 to April 16 were used as testing set to evaluate the performance of the model. Remaining data (April 17-30) were used as cross validation data set to prevent over training of the model. Figure 1 shows the results of ANN model in testing phase for the first simulation, and also the results of the second simulation is shown in figure 2. In these figures H(1) and H(2) indicate the hydrographs estimated by hydrodynamic model for the outlet of RCEW in first and the second simulations respectively.

6 Optimised results

The combination of neural network model and hydrodynamic model gave a clear improvement in the prediction results. It seems that the combination of these two techniques uses the benefits of both methods and shows good performance (figure 1). Also the results of hydrodynamic modelling affected by the suspension of the Macks Greek flow gauging has been improved appropriately (Figure 2).

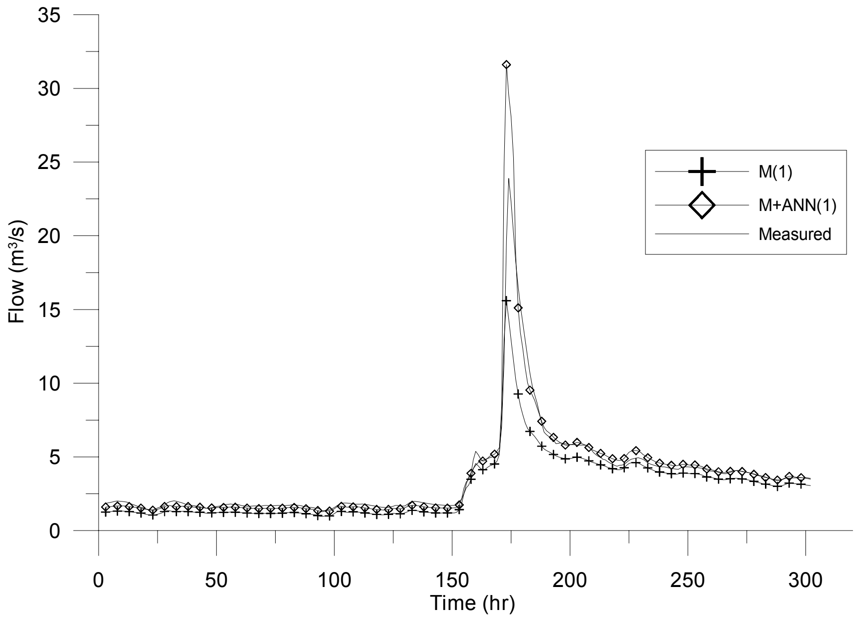

Figure 1 Estimated flow for the outlet of the RCEW using hydrodynamic model (MIKE11) and neural network based optimised values versus measured values (test 1).

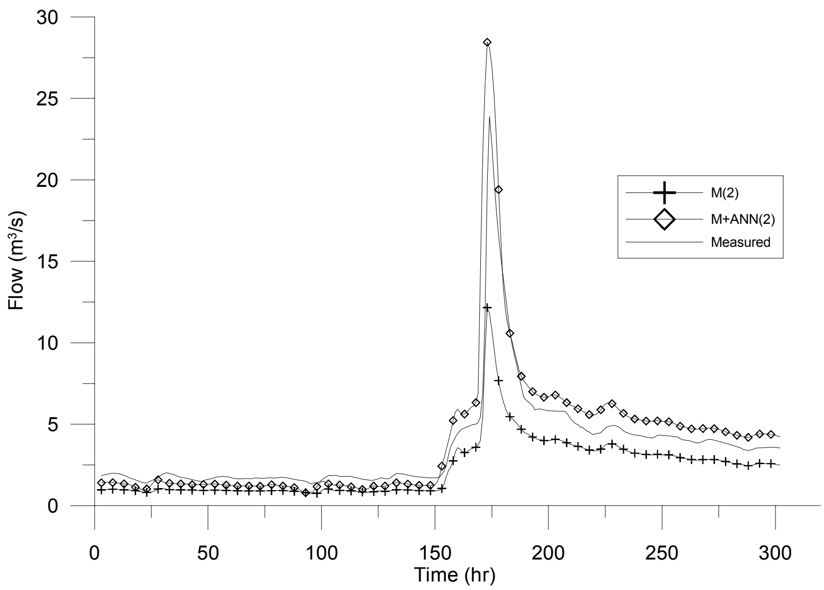

Figure 2 Estimated flow for the outlet of the RCEW using hydrodynamic model (MIKE11) and neural network based optimised values versus measured values (test 2).

To evaluate the accuracy of the results statistically, two criteria were used: coefficient of efficiency, R^2, and root mean square error, rmse. Values of these parameters for the results of study are shown in table 1.

Table 1 Coefficient of efficiency (R^2) and root mean square error (rmse) of the results before and after optimisation

	Initial results (hydrodynamics results)		Optimised results (hydrodynamics +ANN)	
	first simulation	Second simulation	first simulation	Second simulation
R^2	0.788	0.576	0.876	0.767
rmse	0.936	1.324	0.718	0.980

7 Conclusions

The application of artificial neural networks for correction of the outputs of a 1D hydrodynamic flow model in the Reynolds Creek Experimental Watershed, southwest Idaho, has been investigated in this study. After applying a hydrodynamic model using the MIKE11 software, the error of the model was predicted using a three-layer feedforward neural network model to optimise the outputs. This gave a significant improvement in the results. For the first simulation (before suspension of a gauging site in the catchment), the coefficient of efficiency, R^2, increased from 0.788 to 0.876. The root mean square error (rmse) improved from 0.936 m^3/s to 0.718 m^3/s. For the second simulation (after suspension), neural networks improved the results by increasing R^2 from 0.576 to 0.767 and decreasing rmse from 1.324 to 0.980.

Acknowledgement

The authors are grateful to the USDS-ARS Northwest Watershed Research Centre for the use of research data and related reports.

References

Danish Hydraulic Institute, *DHI Software 2000b*, November 2000.
Dastorani M. T. and N. G. Wright, *Application of artificial neural networks for ungauged catchments flood prediction,* presented to the Floodplain Management Association 2001 conference, March 12-16, San Diego, USA.
Hanson C. L., D. Marks and S. S. van Vactor, *Climate monitoring at the Reynolds Creek Experimental Watershed, Idaho, USA*, ARS technical bulletin NWRC-2000-6.
Hanson C. L., *Precipitation monitoring at the Reynolds Creek Experimental Watershed, Idaho, USA*, ARS technical bulletin NWRC-2000-4.
Marks D., K. R. Cooley, D. C. Robertson and A. Winstral, *Snow monitoring at the Reynolds Creek Experimental Watershed, Idaho, USA*, ARS technical bulletin NWRC-2000-5.
NeuroDimensions, *NeuroSolutions*, www.nd.com, 2001.
Pierson F. B., C. W. Slaughter and Z. K. Cram, *Monitoring discharge and suspended sediment, Reynolds Creek Experimental Watershed, Idaho, USA*, ARS technical bulletin NWRC-2000-8.
Seyfried M. S., R C. Harris, D. Marks and B. Jacob, *A geographic database for watershed research, Reynolds Creek Experimental Watershed, Idaho, USA*, ARS technical bulletin NWRC-2000-3.
Slaughter C. W., D. Marks, G. N. Flerchinger, S. S. van Vactor and M. Burgess, *Research data collection at the Reynolds Creek Experimental Watershed, Idaho, USA*, ARS technical bulletin NWRC-2000-7.
USDA-ARS Northwest Watershed Research Center anonymous ftp site: *ftp.nwrc.ars.usda.gov*.
Wright N. G. and M. T. Dastorani, *Effects of river basin classification on Artificial Neural Networks based ungauged catchment flood prediction*, in the Proceedings of the International Symposium on Environmental Hydraulics, December 5-8, 2001, Phoenix, USA.
Wright N.G., N. T. Dastorani, P. Goodwin and C. W. Slaughter, Using artificial neural networks for optimisation of hydraulic river flow modelling results, to be presented to the international conference of river flow 2002, September 4-6, 2002, Louvain-al-Neuve, Belgium.