Application of the combination of hydrodynamic and artificial neural network approaches for river peak flow prediction

Mohammad T. Dastorani[1] *& Nigel G. Wright*[2]

The application of artificial neural networks for the correction of the outputs of a 1D hydrodynamic flow model in a semi arid catchment has been investigated in this study. A hydrodynamic model was constructed to predict flow at the outlet using time series data from upstream gauging sites as boundary conditions the results was not close enough to the actual values. Then the model was replaced by an ANN model but the results were not desirable. Finally 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. Due to suspension of flow gauging in one of the upstream sites, there is no data for this site for the last decade. To evaluate the adaption of models with this problem, all simulations were repeated but without data from the suspended site. A combination of these two techniques produced outputs that were more accurate than the results of the models individually

Introduction

Issues like flooding and associated damage, droughts and water shortages, and compromised water quality are major concerns facing communities worldwide. Understanding these issues requires that the hydrologists have access to high quality long-term data sets to be able to make reliable predictions about future conditions. Completion of hydrologic process research, ecosystem analysis, model development, calibration and validation will help to manage and protect the world's water resources and to mitigate the hazards of an unforeseen future.

In a flood prediction system, especially for large river catchments, a combination of rain fall-runoff and routing models may be used. A rainfall-runoff model is normally used for tributaries while a hydrodynamic or routing model for the main river reaches. The computational models used to predict river floods are in most cases one-dimensional. River floods are normally gradually varied unsteady flows and so a time-dependent simulation is required.

In this research, after consideration of the study area and the most important factors such as climate conditions, precipitation characteristics, flow regime and data gauging network in the Reynolds Creek Experimental Watershed (RCEW), modelling works were completed. Models were developed using hydrodynamic and artificial neural networks separately to predict flow at the outlet of the catchment, and the performances were evaluated. Then a MLP neural network model was adopted to optimise the outputs of the hydrodynamic modelling procedure by prediction of the error produced by the hydrodynamic model.

Study area

The study area for this part of the research is Reynolds Creeks Experimental Watershed (RCEW) southwest Idaho (USA), a typical intermountain region of the western United States.

[1]School of Civil Engineering, University of Nottingham, Nottingham NG7 2RD, UK.
Email: evxmtd@nottingham.ac.uk
[2]School of Civil Engineering, University of Nottingham, Nottingham NG7 2RD, UK.
Email: nigel.wright@nottingham.ac.uk

It is located in Owyhee Mountains of south-western Idaho, about 80 km south-west of Boise with 239 km^2 drainage area. The main stream flows from south to north in the Owyhee mountains at an elevation exceeding 2200m.

The topography of the study area is generally rugged except in the broad valley floor in the north-central part of the watershed. It ranges from a broad, flat alluvial valley to steep, rugged mountain slopes. The elevation ranges over 1100m through the catchment, resulting in a strong climatic gradient. The lower boundary of the catchment is determined by the outlet weir location, which is near the head of a small canyon through which Reynolds Creek flows before entering the Snake River about 12 km to the north. The lowest elevation on the watershed is 1101 m above the sea level and the highest elevation is 2241m at the southern boundary of the catchment. The eastern boundary rises to about 1525m and the western to 830 m above sea level. The climate of the RCEW and more localised distribution of soils and vegetation are largely controlled by the elevation and local topography. The catchment's main perennial stream flow is generated at the highest elevations in the south and north west where deep and late lying snow packs are the main source for water. Local slope and aspect strongly influence the hydrology of the catchment by controlling the incoming solar radiation and snow deposition patterns (Seyfried et al; 2000).

Quality controlled and validated hourly stream flow data sets are available for 13 sites since 1963 (or for subset of that time for some sites) (Pierson et al. 2000). Figure 1 shows the location of all flow gauging stations in the Reynolds Creek Experimental Watershed. One of the gauging stations (Macks Creek) was established early in the 1960s and the measuring of flow in a main tributary was suspended in 1991.

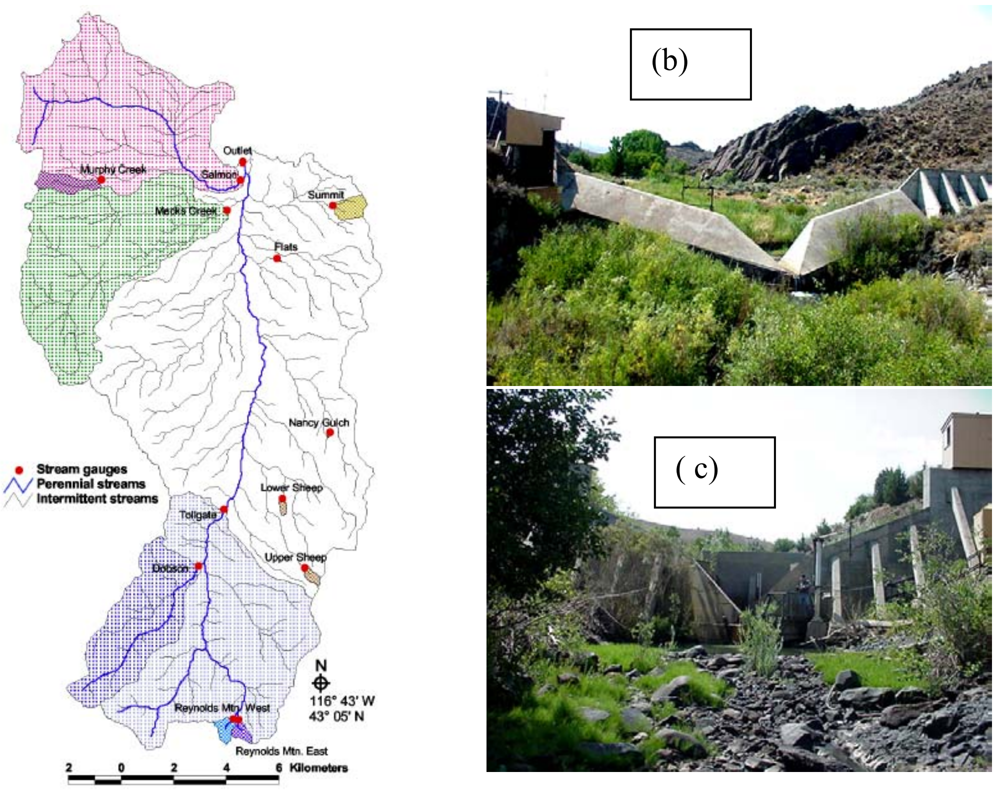

Figure 1 (a) Watershed boundaries, stream network and weir locations in the RCEW (after Pierson et al; 2000), (b) Tollgate weir, the upstream end of the modelled reach and (c) Outlet weir, downstream end of the reach in RCEW.

In this study the Reynolds Creek main river reach between the Tollgate weir upstream and the Outlet weir downstream has been modelled. This reach is a perennial stream with 14.082 kilometres length of stream course (the detailed surveyed thalweg length has been calculated as 17.073 kilometres). The downstream end of the reach is the Outlet weir, which is also the outlet point of the RCEW (figure 1b). The upstream end of the reach is limited to the Tollgate weir (figure 1c). The longitudinal slope of the modelled water way in upper parts of the reach is relatively greater than in lower parts. A total of 53 cross sections have been drawn by surveying the distance between two weirs to provide the data required for calibration of the models.

Flow prediction by hydrodynamic model

The MIKE 11 package (from Danish Hydraulics Institute) was used to construct 1D hydrodynamic model for the reach between the Tollgate and Outlet weirs to predict flow at the Outlet for a flood wave occurring in the period of February-April 1982. Flow data were entered into the model from Tollgate, Macks Creek and Salmon Creek (three gauging stations upstream side of the outlet weir) as the boundary conditions. These three gauging sites which drain three main sub-catchments can be used to model main river reach, while all other gauging sites are located in headwater tributaries of the sub-catchments and not delivering flow directly to the main river. 53 cross sections with associated photographs of the main channel and flood plain in each cross section location were used to define geometry and roughness (Manning's n) to the model. A Q-H relationship at the outlet weir section was used as the downstream boundary of the model. Initially the hydrodynamic model was just used to predict flow at the Outlet point using data from the three sites mentioned above. Figure 2 shows the related hydrograph for March 11 to April 16 predicted by the model against the measured values. As the figure shows, the estimated values are not close enough to the measured values especially for the peak flow. In comparison wih the actual hydrograph the model underestimates the outputs. As a result the output of the model at this stage may not be accurate enough for practical applications.

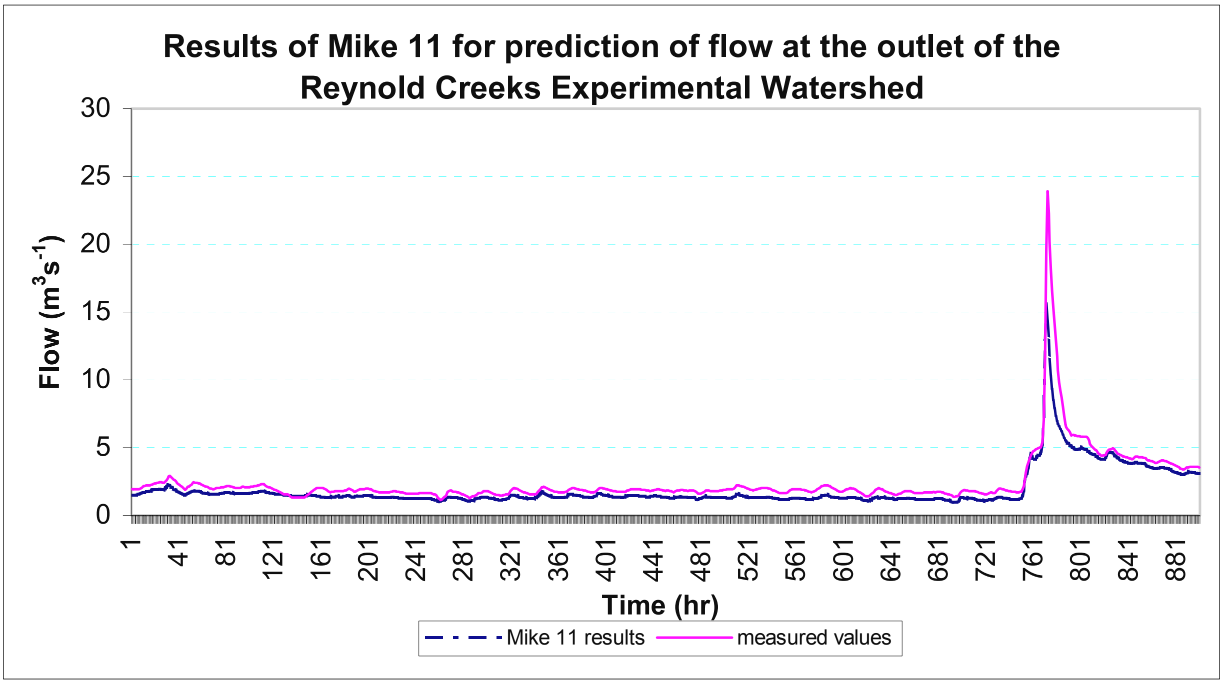

Figure 2 Estimated flow for the outlet of the Reynolds Creek Experimental Watershed using MIKE 11 versus the measured values (using all three upstream sites).

As mentioned earlier the operation of Macks Creek has been suspended since 1991. The second step in this part of the research was to evaluate how this gauging site suspension affects the hydrodynamic modelling and how the technique of artificial neural networks might be able to bridge this gap in order to improve the results of the hydrodynamic model affected by this shortage of measured data. Therefore, another simulation was carried out but this time using data from only two stations where gauging has been continued (Tollgate and Salmon Creek). The estimated flow hydrograph of this simulation is compared to the measured values in Figure 3 (the hydrograph is for the period March 11 to April 16).

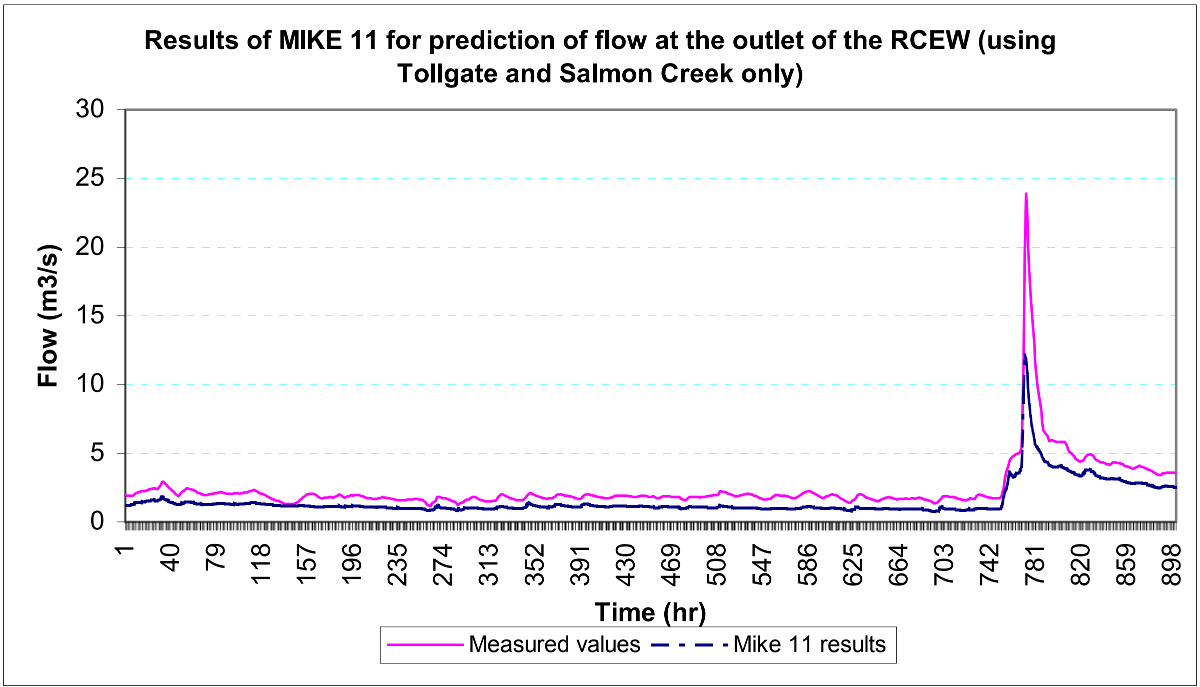

Figure 3 Estimated flow for the outlet of the Reynolds Creek Experimental Watershed using MIKE 11 versus the measured values (using Tollgate and Salmon Creek only).

As can be seen from the figure results are not satisfactory and much worse than the first simulation. This clearly shows the effect of the absence of the Macks Creek as a source of data for the model.

Flow prediction by neural network model

A neural network model was developed and used to predict flow at the outlet point of the RCEW by the same input data as used in hydrodynamic model in previous section. In other words, for the same purpose and same data the hydrodynamic model was replaced by a neural network model. The artificial neural network architecture used was a three-layer perceptron feedforward (MLP) network. Although different architectures of neural networks were tested to be used for this purpose, The MLP was the most relevant one (it presented the most accurate results. One hidden layer with a tangent hyperbolic transfer function was used in this model, while the function in the output layer was a logistic one. To prevent the model from over training, a small part of data was specified and entered as a cross-validation data set. Data was divided into three parts to use as training, testing and cross validation data sets. 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 was used as a testing set to evaluate the performance of the model. The remaining data (April 17-30) was used as cross validation data set. The period of data specified for testing phase was similar to the period of data used in MIKE 11 to facilitates comparison of the results.

In the first simulation all three upstream gauging sites (Tollgate, Macks Creek and Salmon Creek) were used as sources of data to the model. Figures 4 shows the results obtained from this simulation in testing phase. As in the hydrodynamic modelling, the simulation was repeated here without data from the suspended gauging site (Macks Creek) to see how this technique deals with the problem of data source reduction. Figure 5 shows the outputs of this simulation in the testing phase.

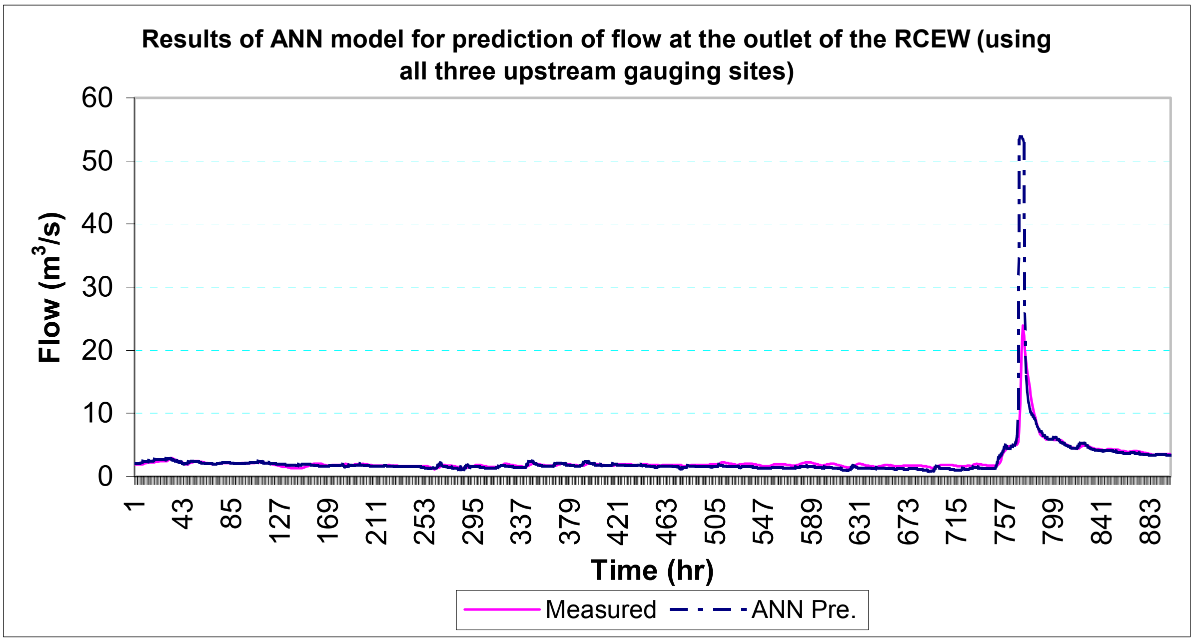

Figure 4 Estimated flow for the outlet of the Reynolds Creek Experimental Watershed using ANN versus measured values in testing phase (all stations).

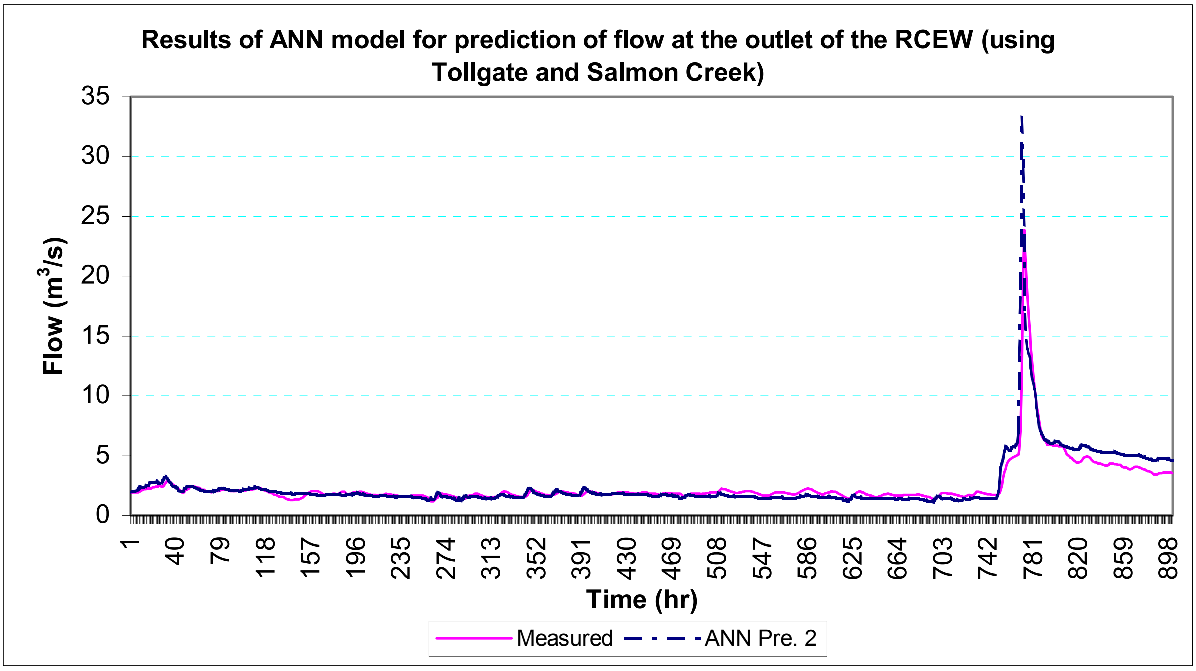

Figure 5 Estimated flow for the outlet of the Reynolds Creek Experimental Watershed using ANN versus measured values in testing phase (no Macks Creek).

Comparing figure 4 figure 5 shows that removal of the Macks Creek data from the ANN model actually caused a significant improvement in the predictions. In addition to the figures the values of R^2 and RMSE (table 1) indicates this improvement quite clearly. This behaviour was not what was expected. However a neural network is a black box model, and depending on the nature of the problem and especially the attitude of data, defines the relationship between input and output, and increasing or decreasing the entered data cannot necessarily improve the performance.

In view of the above it was decided to carry out a simulation after removing a second source of data (Salmon Creek) and therefore use data only from Tollgate, which is located across the main stream and for which the data shows more correlation to those of the outlet. Figure 6 has been drawn using the outputs of the model in this condition against the measured values. This figure shows that removing of Salmon Creek data from the model has affected the results by decreasing the accuracy which is the opposite to what occurred when Macks Creek data was removed. Therefore the neural network model has produced its best possible results in the second simulation (using Tollgate and Salmon Creek to enter data to the model).

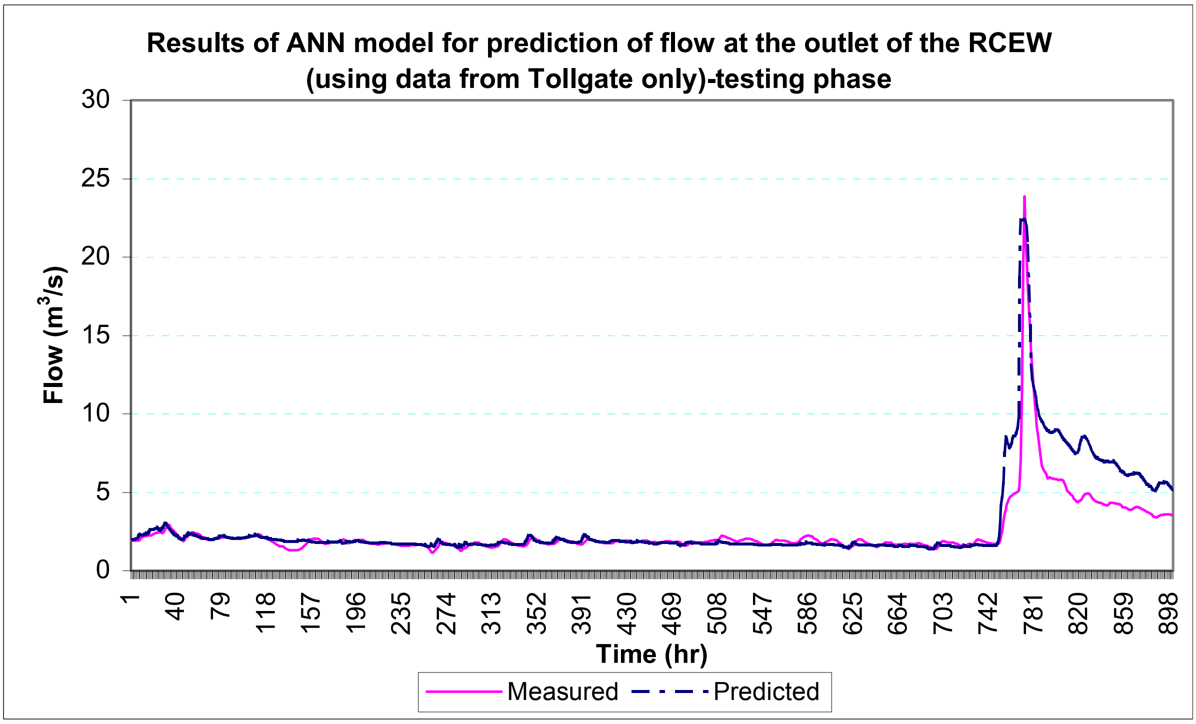

Figure 6 Estimated flow for the outlet of the Reynolds Creek Experimental Watershed using ANN versus measured values in testing phase (no Macks Creek & Salmon Creek).

Combination of ANN and hydrodynamic models

In this part a combination of two techniques (hydrodynamic and ANN) was used to predict flow at the outlet of the RCEW.

ANN model to predict errors

To improve the accuracy of the hydrodynamic model, an artificial neural network was employed to estimate the errors of hydrodynamic modelling results. Different architectures such as Radial basis function, Recurrent network, Time Lag Recurrent network and Multi-layer Perceptron (MLP) were used and the predicted errors were compared to the actual errors. The MLP gave the most accurate results and therefore it was selected as the neural network architecture to be combined with the hydrodynamic model in this part of the research.

The error of the hydrodynamic model estimations was calculated using the following formula:

$$E_p = X_{obs} - X_{est} \qquad (5.4)$$

where E_p is the error of the estimated pattern, X_{obs} is the observed value and X_{est} is the estimated value. The ANN model was trained using Tollgate measured flow data and MIKE 11 outputs for the outlets as input to the ANN model and the difference between MIKE 11 outputs and the measured values for outlet (error of MIKE 11 estimation) as output of the neural network model (E_P).

Hourly flow data for the period of February 1 to April 30, 1982 was used for this simulation (similar to that used in hydrodynamic and ANN models in previous sections). The prediction of error for the first (using all three stations) and second (without Macks Creek) hydrodynamic simulations was carried out separately. The data sets for training, testing and cross validation were the same as in the previous ANN modelling. Figure 7 shows the results of ANN model in this simulation in testing phase.

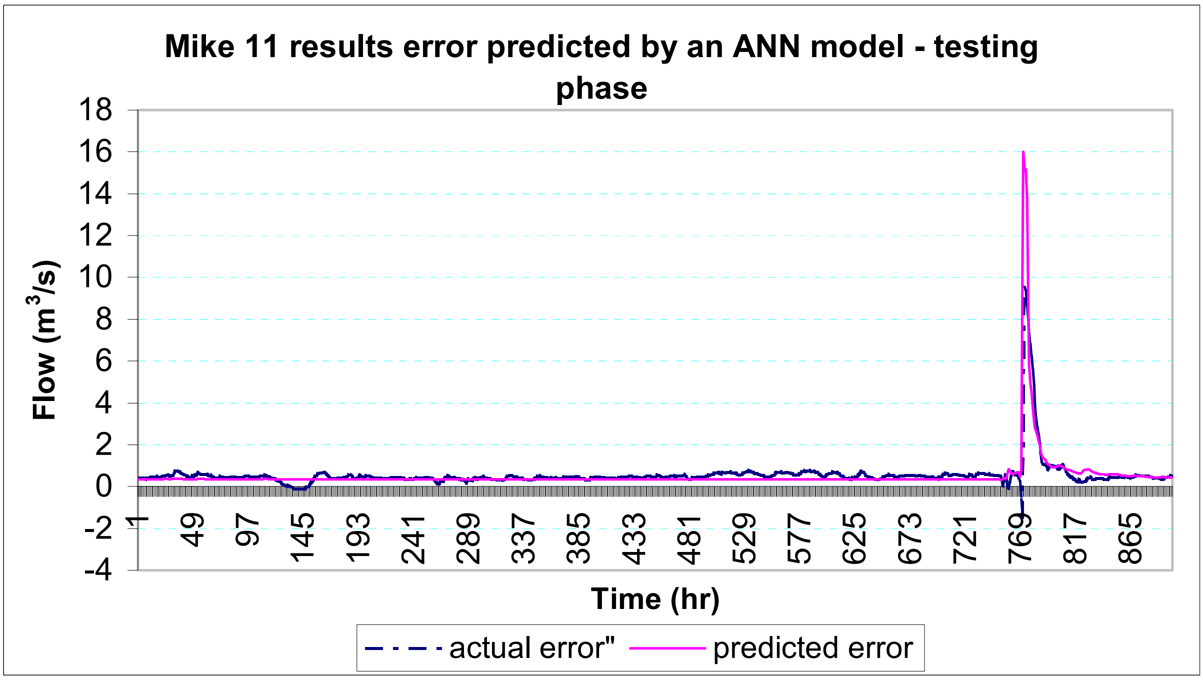

Figure 7 Prediction of the error of MIKE11 outputs for outlet of the Reynolds Creek Experimental Watershed using an ANN model-testing phase.

Optimised results and discussion

The combination of a hydrodynamic model and ANN caused a clear improvement in the results. As figure 7 shows the MLP neural network makes good prediction of errors. The left side of the graph, where the small errors occurred (low flow part of the hydrograph), the predicted line does not follow the small variations of the actual error graph but produces values which are a general average of the error values in this part of the graph. However, when the maximum error occurs (right side of the graph where corresponds to the peak flow period of the flow hydrograph) the predicted line follows the actual line. In this part of the graph the variation of the error is considerable and the closeness of the predicted and measured lines is of importance. It seems that the combination of these two techniques for this specific application uses the potential of both methods (Figure 8). The optimised flow hydrograph of the MIKE11 model for the outlet of the RCEW using ANN after suspension of the Macks Creek gauging site (without using data from this site) is shown in Figure 9. The results have been improved considerably in comparison to figure 3, which shows the results of the hydrodynamic model for the same condition (without data from Macks Creek). The results shown in figure 9 also indicate an improvement in prediction of the peak flow in comparison to the results of the ANN model (figure 5).

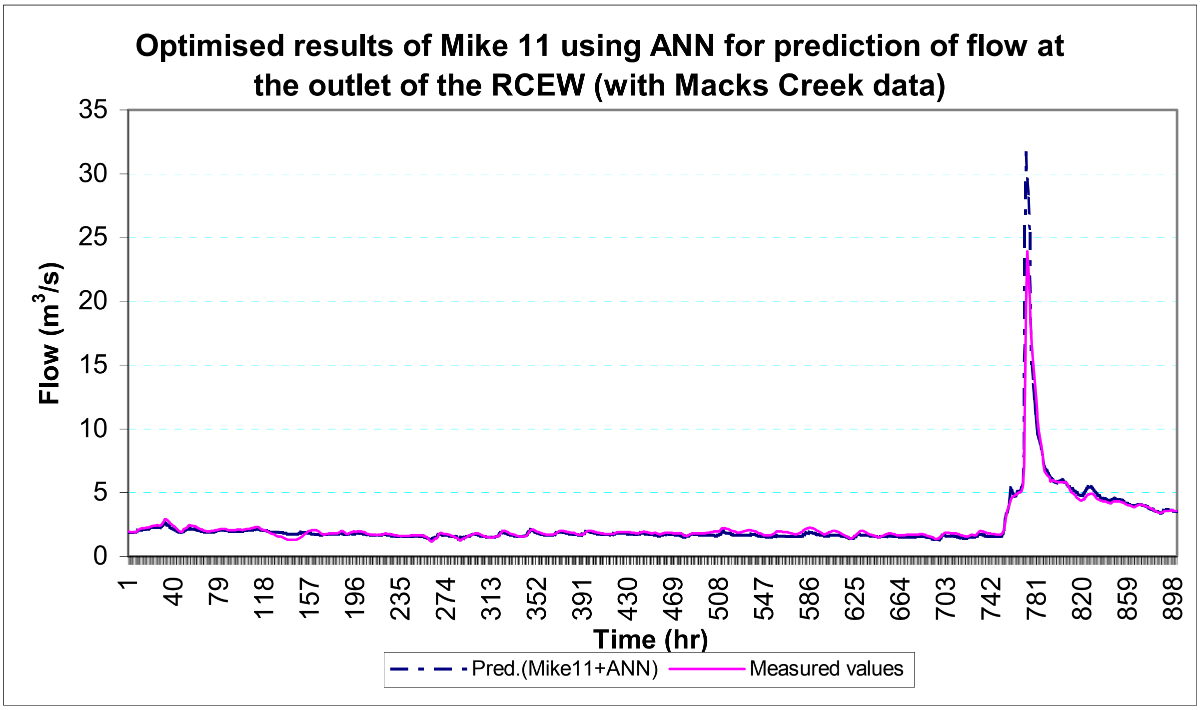

Figure 8 Optimised flow hydrograph of MIKE11 for outlet of the Reynolds Creek Experimental Watershed using ANN (with Macks Creek).

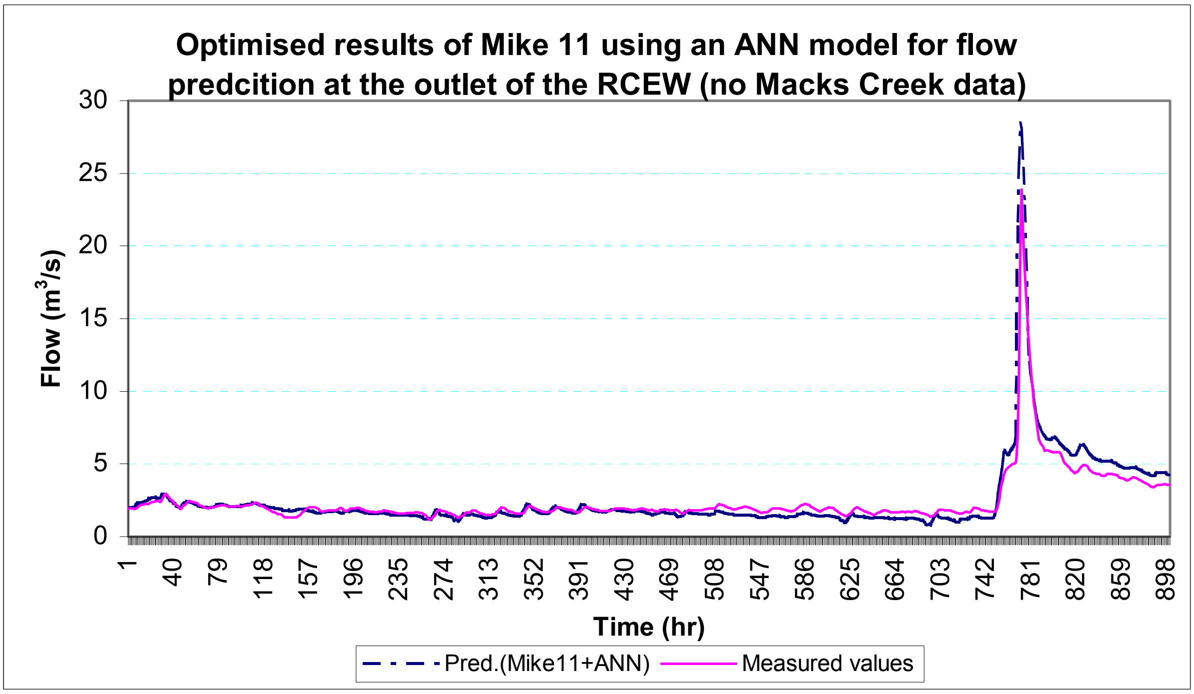

Figure 9 Optimised flow hydrograph of MIKE11 for outlet of the Reynolds Creek Experimental Watershed using ANN (without Macks Creek).

In general, the predicted values for the low flow period (the left side of the hydrographs) seems closer to the actual values compared to the high flow period predictions (the peak flow in right hand side of the hydrographs). For example the ANN model using all three gauging sites (figure 4) performed very well on prediction of the low flow but its output for peak flow period is extremely overestimated.

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

	First simulations (before suspension)			Second simulations (after suspension)		
	MIKE 11	ANN	MIKE11+ANN	MIKE 11	ANN	MIKE11+ANN
R^2	0.6435	0.5432	0.9153		0.8035	0.8725
RMSE	0.9364	2.8143	0.7158	1.3242	1.0813	0.9806

Looking at the table makes the point clear that the improvement of the accuracy of the results produced by the combination of these two models is over those produced by them individually. This is true about all simulations either with or without using data from Macks Creek. For the results of the hydrodynamic model alone, the outputs are generally underestimated. This is mostly due to the absence of a part of the catchment runoff as the boundary condition of the model. The south sector of the catchment has the most precipitation and runoff and most of this flow is measured at the Tollgate site. Flow measured in two other main tributaries (Macks Creek and Salmon Creek) are also considered to the model but there are still several small tributaries draining about 40% of the catchment that have no data to enter to the model. Getting under estimated predictions from the hydrodynamic model is to be expected as part of the catchment is not contributing to the model. The unsatisfactory nature of the results produced by the ANN model alone is not however a simple task to explain. Neural network modelling is a black box method and establishment of an efficient input-output relationship is strongly case and data dependent. The reliability of the outputs is mostly dependent on these relationship established between the inputs and outputs. A strong relationship in the training phase normally gives the most accurate output in the testing phase (as long as the testing data set is in the range of training data and the model is not over trained). The strength of this relationship depends on the correlation between the data series used as inputs and outputs. In most cases this correlation depends on the range, order and nature of data sets rather than closeness of the values or the number of input patterns. For example, correlation between a series such as {1, 2, 3, 4, 5} and {10, 20, 30, 40, 50} is stronger than correlation between a series such as {1, 2, 3, 4, 5} and {9, 21, 29, 42, 52}. Removing the data of a gauging site in the present case removes the flow data of a part of catchment from the hydrodynamic model boundary and causes underestimation of the outputs. However, in the neural network model removing of the Macks Creek data caused better correlation between the combination of Tollgate and Salmon creek data and the outlet data, and finally leads to better performance of ANN model. Even simulation with only one input pattern (Tollgate data) produced results better than the simulation with three input patterns. In modelling with this technique there are no certain rules to set all required parameters to reach desirable results, and trial and error procedure in some aspects of modelling by this technique is the only way to improve the output accuracy.

Replacement of the hydrodynamic model by a neural network model with the same input data produced results completely different from the hydrodynamic model. For the low and normal flow discharge, ANN produced results are quite close to the measured values. However, in prediction of the high flow discharge (flood wave) the outputs are considerably over estimated in contrast to the hydrodynamic model. In contrast to the outputs of the hydrodynamic model, removing the Macks Creek data from the ANN model did not decrease the accuracy of the results but in fact caused a significant improvement of the predictions. In this method quality of data regardless of the size of data set in some cases can change its performance considerably. Improvement of the accuracy by removing Macks Creek data prompted an investigation of removing the second source of data (Salmon Creek) and use data only from Tollgate. However, removing of the Salmon Creek data from the model has affected the results by decreasing of the accuracy in contrast with what occurred when Marks Creek data was removed. Therefore the neural network model has produced its best possible results in the second simulation (using Tollgate and Salmon Creek to enter data to the model).

Conclusion

The Neural Network technique is an appropriate predictor of the error in hydrodynamic models. For practical purposes the improvement of predictions is crucially important as it strongly affects the costs and risk of projects. Coupling of the existing hydrodynamic methods with new machine learning tools such as artificial neural networks seems an important step forward to more reliable prediction of river flow.

References

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.

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.

USDA-ARS Northwest Watershed Research Center anonymous ftp site: *ftp.nwrc.ars.usda.gov*.

Wright N.G., N. T. Dastorani, P. Goodwin and C. W. Slaughter, Using artificial neural networks for optimisation of hydraulic river flow modelling results, International Conference of River Flow 2002, September 4-6, 2002, Louvain-al-Neuve, Belgium.