East African Scholars Journal of Engineering and Computer Sciences
Abbreviated Key Title: East African Scholars J Eng Comput Sci
ISSN 2617-4480 (Print) | ISSN 2663-0346 (Online) |
Published By East African Scholars Publisher, Kenya
Volume-2 | Issue-10 | Oct-2019 |

OPEN ACCESS

Review Article

A Flood Forecasting Model Based on Artificial Neural Network

Nghiem Van Tinh
Thai Nguyen University of Technology – Thai Nguyen University, ThaiNguyen, Vietnam

*Corresponding Author
Nghiem Van Tinh

Abstract: In the real time, floods have been a major cause of loss of life and property. Methods of prediction and mitigation range from human observers to sophisticated surveys and statistical analysis of climatic data. This paper presents a flood forecasting model to predict flood in rivers based on Artificial Neural Network (ANN). The river system chosen for the research was the Big Thompson River, located in North-central Colorado, United States of America. The study show that the forecast results in term of prediction accuracy of greater than 96% in +/-100 cubic feet per minute range. The average error of the predictions was less than 16 cubic feet per minute. To further validate the model's predictive capability, a multiple regression analysis was done on the same data. The Neural Network's predictions exceeded those of the multiple regression analysis by significant margins in all measurement criteria.
Keywords: Forecasting, regression analysis, neural networks, flood.

1. INTRODUCTION

Natural disasters such as flood and tropical cyclones are regarded to be caused by extreme weather conditions as well as changes in global and regional climate. United States of America and other countries are faced with environmental and ecological challenges particularly in view of the impact of climate change. These include the occurrence of natural disasters such as fire, floods, tropical storms, major accidents, drought, epidemic diseases and food shortage. What is needed in flood forecasting is a system that can be continuously updated without the costly and laborious resurveying that is the norm in floodplain delineation. In recent years, many published papers have shown the results of research on Neural Networks (NN) and their applications in solving problems of control, prediction, and classification in industry, environmental sciences, and meteorology (French, Krajewski, & Cuykendall, 1992); (Boznar, M., & Mlakar, 1993); (Aussem, Murtagh, & M., 1995); (Blankert, 1994); (Ekert, Cattani, & Ambuhl, 1996). Schultz (1996) demonstrated and compared three models of rainfall-runoff models using remote-sensing applications as input. The first model was a mathematical model which demonstrated the ability to reconstruct monthly river runoff volumes based on infrared data obtained by the Meteosat geostationary satellite. The second model computes flood hydrographs from a distributed system rainfall/runoff model. Lee and Singh (1999) presented a Tank Model using a Kalman Filter to model rainfall-runoff in a river basin in Korea. Choy and Chan (2003) used an associative memory network with a radial basis functions based on the support vectors of the support vector machine to model river discharges and rainfall on the Fuji River. Another study by Neary, Habib, and Fleming (2004) used the Hydrologic Modeling System developed by the Hydrologic Engineering Center. The model, commonly referred to as HMS-HEC, is widely used for hydrologic modeling, forecasting, and water budget studies. This paper is an effort to demonstrate the potential use, by a layperson, of a commercially available NN to predict stream flow and probability of flooding in a specific area. In addition, a comparison was made between a NN model and a multiple-linear regression model.

The rest of this paper is organized as follows. Section 2 introduces neural networks model. Section 3 applies neural networks for predicting flood events in the Big Thompson River, located in North-central Colorado, United States of America and conclusions are presented in Section 4.

2. NEURAL NETWORK MODEL

Neural networks have been successfully applied to the forecasting of different applications as

Quick Response Code

Journal homepage:
http://www.easpublisher.com/easjecs/

Article History
Received: 13.09.2019
Accepted: 26.09.2019
Published: 10.10.2019

DOI: 10.36349/easjecs.2019.v02i10.003

credit ratings (Kumar & Bhattacharya, 2006), Dow Jones forecasting (Kanas, 2001), customer satisfaction analysis (Gronholdt & Martensen, 2005), stock ranking (Refenes, Azema-Barac & Zapranis, 1993), and tourism demand (Law, 2000; Palmer, Montaño & Sesé, 2006). The nonlinear structures of neural networks have been very useful in forecasting and they have been shown to discover nonlinear relationships among the observations (Donaldson & Kamstra, 1996; Indro, Jiang, Patuwo & Zhang, 1999).

Advantages of using NNs include the following:

- ✓ A priori knowledge of the underlying process is not required.
- ✓ Existing complex relationships among the various aspects of the process under investigation need not be recognized.
- ✓ Solution conditions, such as those required by standard optimization or statistical models, are not preset.
- ✓ Constraints and a priori solution structures are neither assumed nor enforced (French *et al.*, 1992).

In this section, the topology of a neural network is specified by the number of layers, the number of units per layer and the weighted connections among all the units. These types of layers are the Input layer, the Hidden layer (of which there may be none too many), and the Output layer (Fu, 1994) as shown in Figure 1. In a feed-forward network, data flows as indicated by the arrows, from the Input to the Output layer.

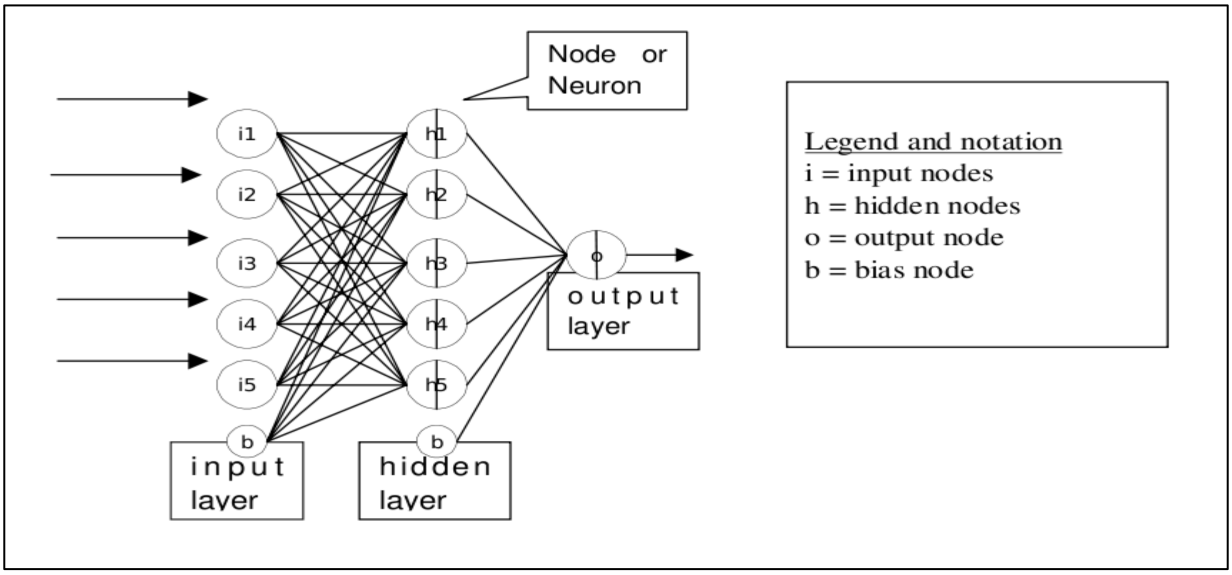

Figure 1. A neural network structure [Fu, 1994]

The Input layer receives input signals or data from the external world and a node in this layer is called an Input unit. These units represent and encode the data or signal pattern presented to the network for processing (Fu, 1994, Kumar *et al.*, 2004).The layer following the Input layer is the Hidden layer, and the nodes in this layer are called Hidden units. The Hidden layer can consist of one or more layers of neurons with the succeeding layers receiving input from preceding layers in feed-forward architecture. The Output layer is the final layer of the network, and the nodes in this layer are called Output units. These units represent encoded concepts (or values) for the training application under consideration..

3. APPLY NEURAL NETWORK FOR FORCASTING FLOOD EVENTS

In this study, the Ward Systems Neural Shell Predictor (http://www.wardsystems.com) is applied to model rainfall/snowmelt-runoff relationship using observed data from the Big Thompson watershed located in North-central Colorado (http://lwf.ncdc.noaa.gov/oa/ncdc.html). For this study of the Big Thompson Watershed, six climatic observation stations were used for the input variables. For the purposes of building a model to demonstrate the feasibility of using the commercially available NN, all six stations' data were used for the independent variables. The following table outlines the steps taken in creating forecasting model which is shown in Table 1.

Table 1: Steps in the use of Neural Networks

Step—problem input	Activities	Definitions and comments
Organizing the Data Buffering the Data	Problem Input. Cleansing the data	Elimination of 'No report' days.
Build the neural network	**Training**	
-Select Strategy - Selecting training set - Selecting the run set Train the Network		Establish the nodes, paths and weights for nodes and paths. Use multiple runs to smooth the input error terms and optimize the desired characteristic (Correlation or MSE,) Smoothing factors (weights) are the only adjustable variables in the Genetic model.
Apply the neural network	**Activation**	
Run the model using hold-out set of data		Back propagate to adjust the weights and eliminate smoothed inputs.
Run another iteration using hold-out set		Testing the model.
Post network and	**Problem output**	
Problem Output	File export, data examination, printouts.	Organize and evaluate efficiency of the model

3.1. *Evaluation of Model Reliability*

Two indicators are used to show the performance of the forecasting model. The network performance statistic known as R-Squared, or the coefficient of multiple determination, is a statistical indicator usually applied to multiple regression analysis. It compares the accuracy of the model to the accuracy of a trivial benchmark model wherein the prediction is just the average of all of the example output values.

➢ ***Experimental Results For The First Run***

The initial run of the data that did not include the previous day's stream flow and the Lake Estes discharge. It resulted in promising but not particularly good results. The following charts demonstrate the initial runs.

These charts depict the actual values versus the predicted *cfm* flow values using data from the five climatic gauging stations. The measuring stations are Drake and Loveland. As one can see, there is a definite correlation between the input data and the resulting values. However, the extreme values are very poorly predicted.

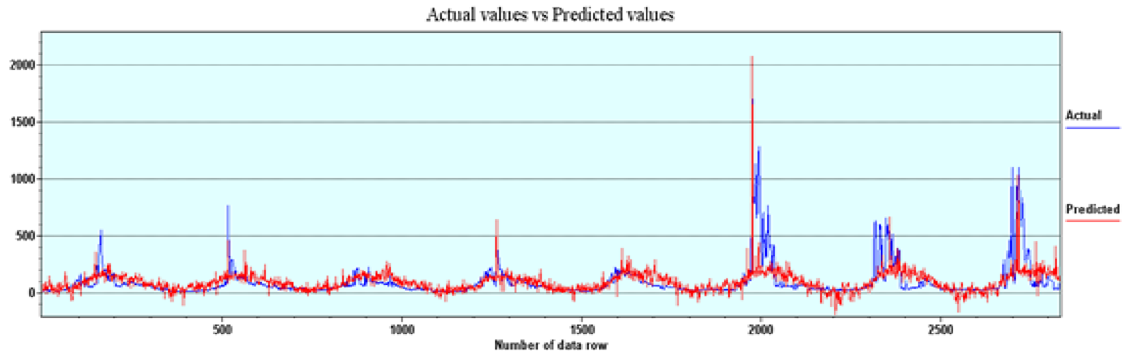

Figure 2. Drake, Initial Run Actual vs. Predicted Values

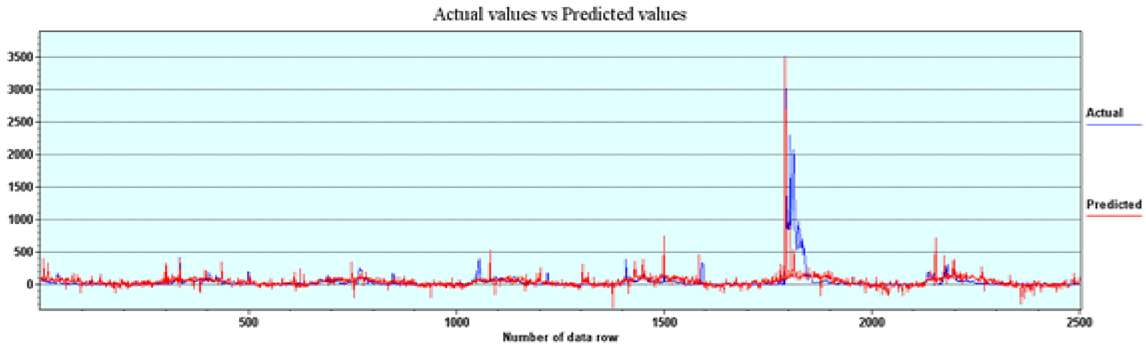

Figure 3. Loveland, Initial run Actual vs. Predicted Values

The R-Squared results are depicted in the graph below. The R-Squared started at a value of approximately 0.24 and improved over the addition of 80 hidden neurons to an approximate 36 value. While promising, the results were not good enough to use as a predictive program.

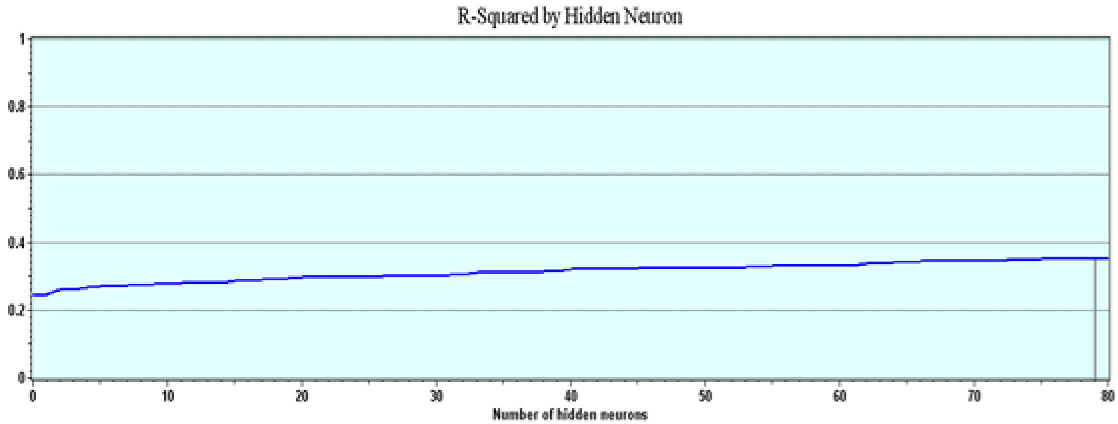

Figure 4. Drake, Initial Run R-Squared

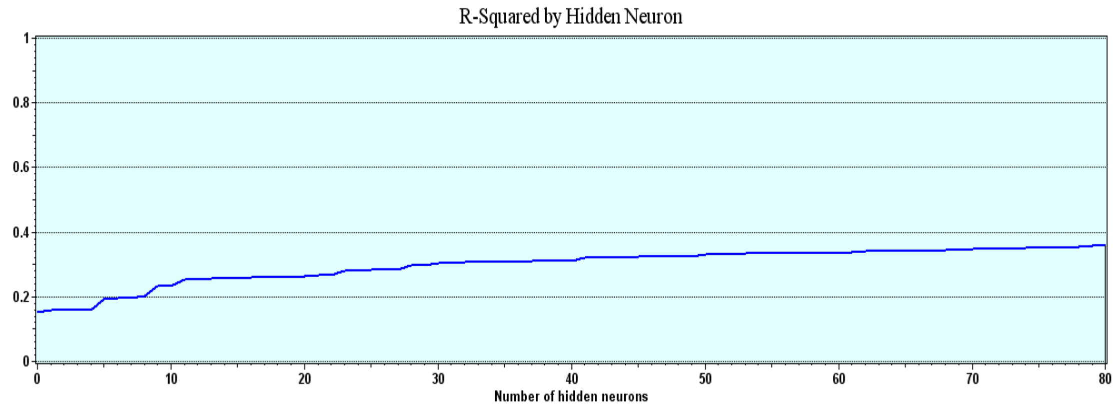

Figure 5. Loveland, Initial Run R-Squared

➢ ***Experimental results for the second run***

The second run was initiated by adding outflow data from the main power plant dam located at Lake Estes on the upper Big Thompson River. This is the controlling dam on the Big Thompson River, which is situated above the two measuring stations that this study uses for the model. All inputs are identical to the first run.

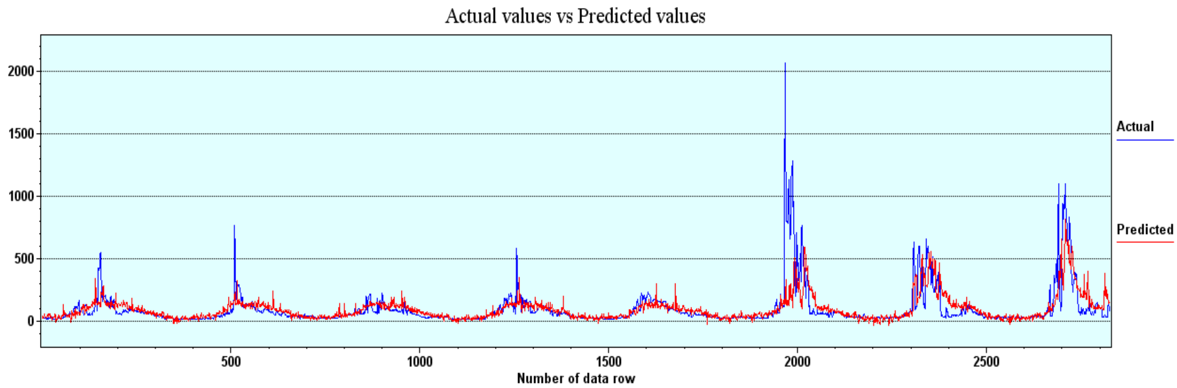

Figure 6. Drake, Second Run, Actual verses Predicted

The actual value verses predicted values for the Drake measuring station and the Loveland measuring station both show definite improvement over the previous run. This run, with the outflow from Lake Estes, still is rather poor on predicting the extreme values associated with flooding events and as such are not adequate.

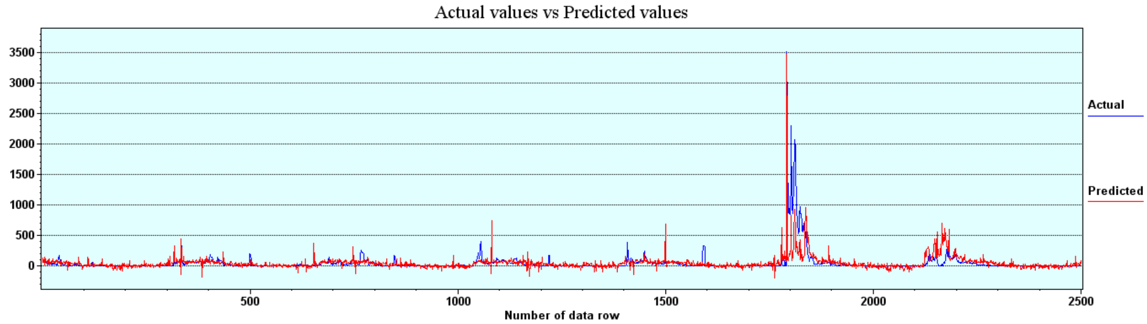

Figure 7. Loveland, Second Run, Actual verses Predicted

The R-Squared value for this run at the Drake measuring station started just above .4 and did not improve through the addition of 80 hidden neurons. The R-Squared values for the Loveland measuring station started just above 0.24 and improved over the addition of 80 neurons to a value of 0.4600. Both stations showed significant improvement for the R-Squared values over the values from the first run.

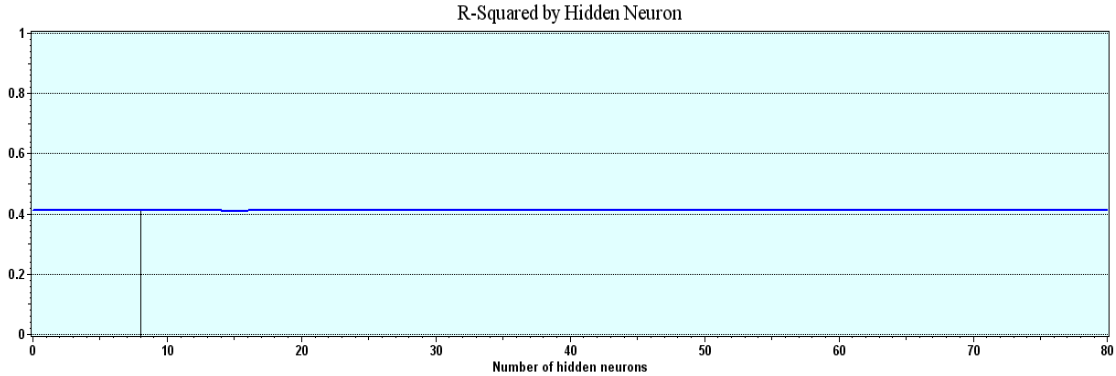

Figure 8. Drake, Second Run, R-Squared

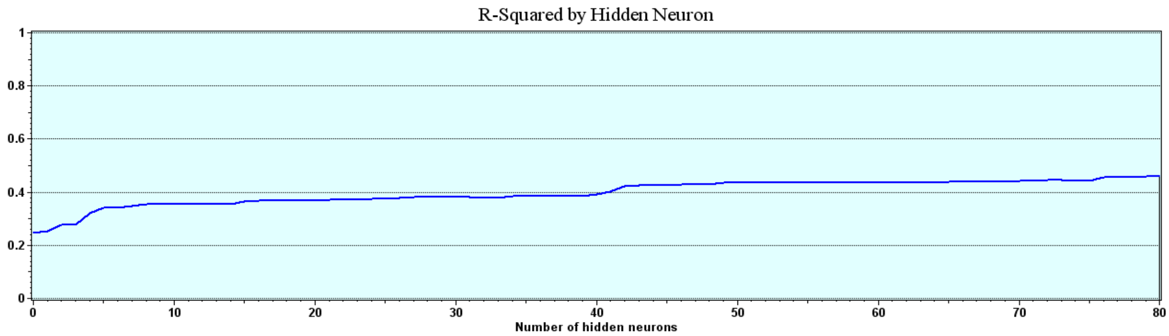

Figure 9. Loveland, Second Run, R-Squared

➢ ***Experimental results for the final run***

The final run was initiated after a major breakthrough occurred in this research, which was the finding and implementing a technique used by (Hsu *et al.*, 1996). This technique demonstrated that results were significantly improved by adding the previous day's stream-flow or stage-level input with the other data. The same inputs are used in this run of data as were used in the two previous models. The new input for this data run is the previous day's flow at the Drake and Loveland measuring stations, respectfully.

The Actual versus Predicted results for both the Drake and the Loveland measuring stations are greatly improved in this final model as demonstrated by the charts below and the following statistical analysis. One extreme event occurred during this time period that was well out of the range of data available and was not adequately predicted by this NN. It is well known that a NN cannot predict an event that it has never seen before in the training data. There was no repeat of the magnitude of this event during the time period under study.

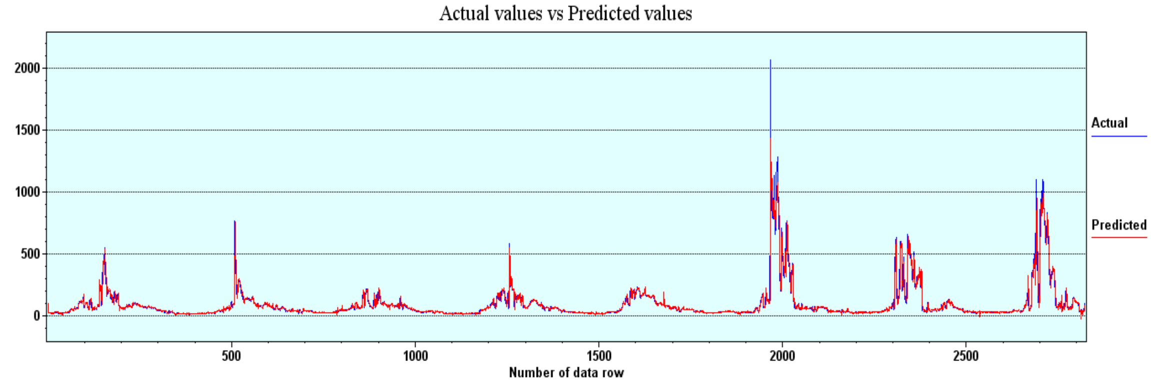

Figure 10. Drake Final Model, Actual versus Predicted

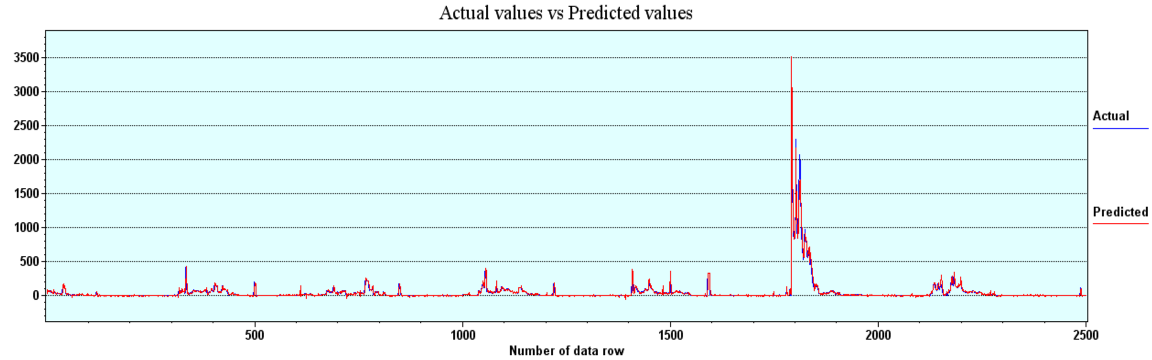

Figure 11. Loveland, Final Model, Actual versus Predicted

R-Squared for this run improved greatly over the first two models for both measuring stations. The Drake measuring station results for R-Squared started at just under 0.90 and improved slightly over the addition of 80 hidden neurons to a value of 0.9091.The R-Squared results for Loveland started at about .86 and improved over the run of data to a value of .9671.

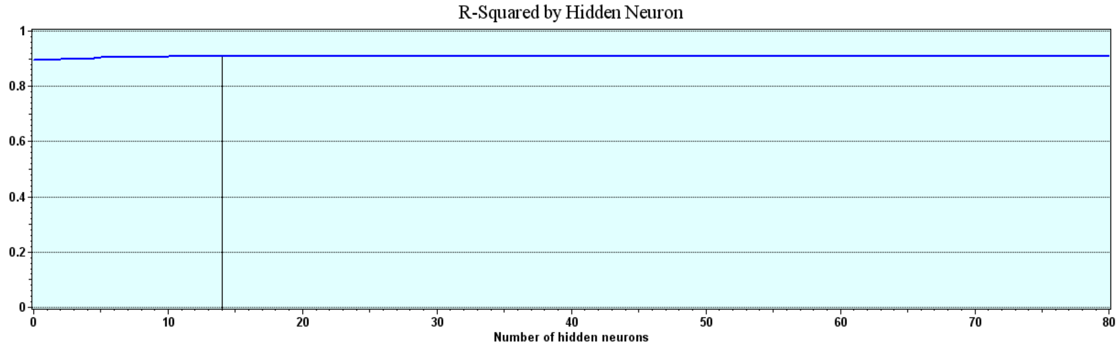

Figure 12. Drake, Final Model, R-Squared.

.

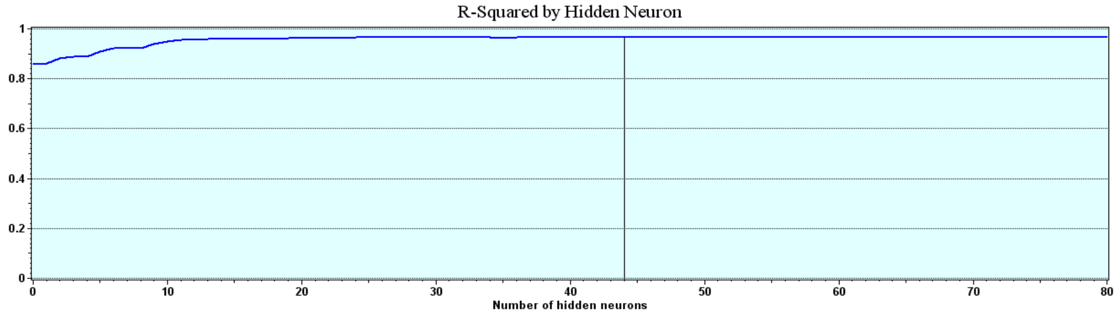

Figure 13. Loveland, Final Model, R-Squared.

The Average Error for the final model improved dramatically over the results of the first two models. Both the Drake measuring station and the Loveland measuring station showed very tight average errors. The Average Error for the Drake Measuring station started the run at about 15.7 cubic feet per minute and decreased over the run to a final value of 15.24 cubic feet per minute.

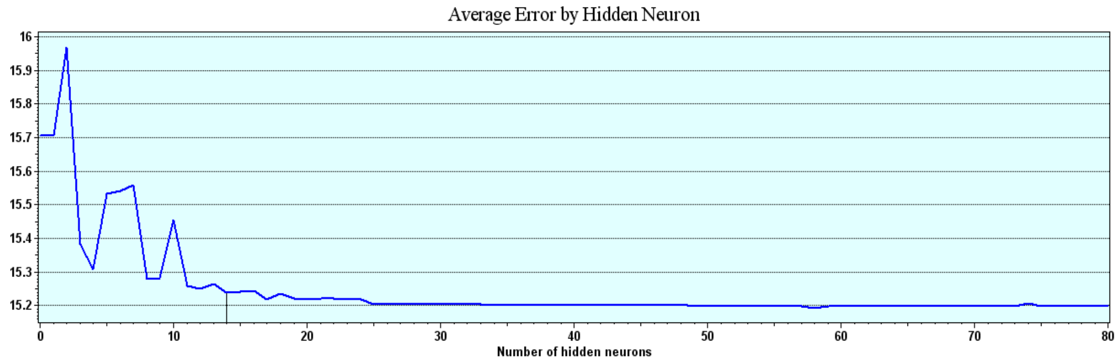

Figure 14. Drake, Final Model, Average Error.

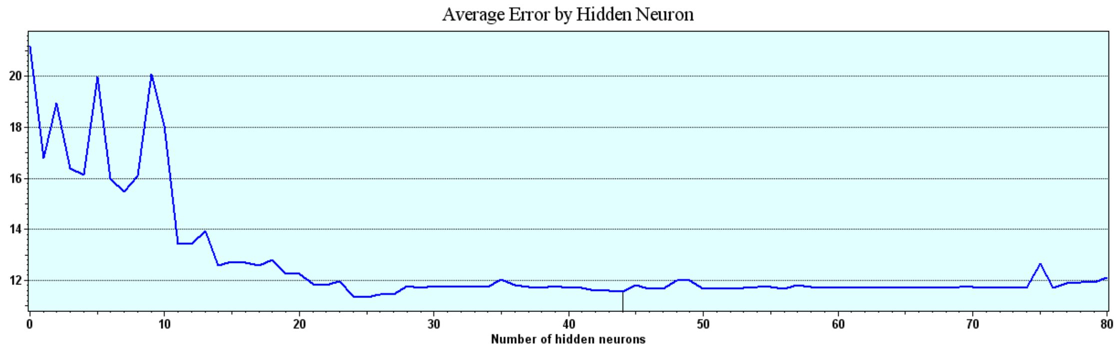

Figure 15. Loveland, Final Model, Average Error

The Average Error for the Loveland measuring station started the run at about 26 cubic feet per minute and decreased to a value of 11.56 Cubic feet per minute.

3.2. Compared with Multi-linear Regression Model

The following Multi linear regression models were created and provided by Dr. Kadar Mazouz of Florida Atlantic University (Mazouz, 2006). A stepwise multi-linear regression model was generated for both data sets, Drake (Appendix A) and Loveland (Appendix B). Being a multiphase process, it stopped after the seventh model. For the Drake measuring station, it gave an R-square of 0.849, which is less than the 0.9091 R-square the NN Model generated for the Drake Data sets.

For the Loveland, the stepwise Multi-linear regression model was generated in eight iterations. It ran R-square of 0.803, which is less than the 0.9671 R-square generated for the Loveland data using NNs.

The Statistical Measures Of These Models Is Shown In Table 2 As Follows:

Table 2: Statistical analysis of the Neural Network model and Multi – linear regression model

Neural Network model					
	R-squared	**Av.Error**	**corrilation**	**MSE**	**RMSE**
Drake	0.9091	15.24	0.9534	1993.011	44.64
Loveland	0.9671	11.56	0.9834	1016.943	31.89
Multi – linear regression model					
Drake	R-squared	Adj.R square	Std.Error of the Estimate		
Loveland	0.849	0.848	28.1527		
	0.803	0.802	20.76851		

4. CONCLUSION

In this paper, a daily rainfall-runoff model for two flow-measuring stations, Drake and Loveland, on the Big Thompson River in Colorado, was developed using a Ward System NN program called the Neural Shell Predictor. The study attempts to demonstrate the feasibility of using a commercially available NN to accurately predict day-to-day normal flows of a river and to predict extreme flow conditions commonly called flood events. In developing this model, the following topics were addressed: (a) the use of a commercially available NN in the development of the daily rainfall, snowmelt, temperature-runoff process; (b) the evaluation of the reliability of future predictions for this NN program; and (c) the comparison of results of the to a Linear Multiple Regression model developed by Dr. Mazouz.

Although the network trained in this study can only be applied to the Big Thompson River, the guidelines in the selection of the data, training criteria, and the evaluation of the network reliability are based on statistical rules. Therefore, they are independent of the application. These guidelines can be used in any application of NNs to other rivers.

REFERENCES

1. French, M. N., Krajewski, F. W., & Cuykendall, R. R. (1992). Rainfall forecasting in space and time using a neural network. *Journal of Hydrology*, *137*, 1-31.
2. Boznar, M., M., L., & Mlakar, P. (1993). A Neural Network-Based Method for Short-term Predictions of Ambient SO2 Concentrations in Highly Polluted Industrial Areas of Complex Terrain. *Atmospheric Environment*, *27B*, 221-230.
3. Aussem, A., Murtagh, E., & M., S. (1995). Dynamical recurrent neural networks-Towards environmental time series prediction. *International Journal of Neural Systems*, *6*, 145-170.
4. Hsu, K. L., et al. (1996). *A superior training strategy for three-layer feed forward artificial neural networks* (Technique report No. HWR no. 96-030). Tucson: University of Arizona, Department of Hydrology and Water Resources.
5. Ekert, P., Cattani, D., & Ambuhl, J. (1996). Classification of Ensemble Forecasts by Means of an Artificial Neural Network. *Meteorology Applications*, *3*, 169-178.
6. Gronholdt L. & Martensen A. (2005). Analysing customer satisfaction data: A comparison of regression and artificial neural networks. International Journal of Market Research, vol. 47, no. 2, pp, 121-130.
7. Kanas A. (2001). Neural network linear forecasts for stock returns. International Journal of Finance and Economics, vol. 6, no. 3, pp. 245-54.
8. Kumar K. & Bhattacharya S. (2006). Artificial neural network vs. linear discriminant analysis in credit ratings forecast: A comparative study of prediction performances. Review of Accounting & Finance, vol. 5, no. 3, p. 216.
9. (Kumar & Bhattacharya, 2006), Dow Jones forecasting (Kanas, 2001), customer satisfaction analysis (Gronholdt & Martensen, 2005), stock ranking (Refenes, Azema-Barac & Zapranis, 1993), and tourism demand (Law, 2000; Law & Au, 1999; Martin & Witt, 1989; Palmer, Montaño & Sesé, 2006)
10. Law R. (2000). Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting. Tourism Management, vol. 21, pp. 331-340.
11. Neary, V. S., Habib, E., & Fleming, M. (2004). Hydrologic modeling with NEXRAD precipitation in middle Tennessee. *Journal of Hydrologic Engineering*, *9*(5), 339-349.
12. Choy, K. Y., & Chan, C. W. (2003). Modelling of river discharges and rainfall using radial basis function networks based on suppor vector regression. *International Journal of Systems Science, 34*(14-15), 763-773.
13. Palmer A., Montaño J. J., & Sesé A. (2006). Designing an artificial neural network for forecasting tourism time series. Tourism Management, vol. 27, pp. 781-790.
14. Refenes A.N., Azema-Barac M., & Zapranis A.D. (1993). Stock ranking: Neural networks vs. multiple linear regression. IEEE International Conference on Neural Networks, pp. 1419-1426.
15. Fu, L. (1994). Neural Networks in Computer Intelligence. McGraw-Hill, Inc., University of Florida, Gainesville.
16. Kumar, D. N., Raju, K. S., and Sathish, T. (2004). River Flow Forecasting Using Recurrent Neural Networks. Water resources management, Kluwer Academic Publishers, Printed in the Netherland, 18:143–161.
17. Mazouz, K. (2006). Multiple Regression Analysis of the Big Thompson River (pp. 3). Boca Raton: Florida Atlantic University.
18. Lee, Y. H., & Singh, V. P. (1999). Tank Model Using Kalman Filter. Journal of Hydrologic Engineering, 4(4), 344-349.