

Using a neural network model to forecast flood events on the Big Thompson River

W. S. Huffman¹ & A. K. Mazouz²

¹*Nova South-eastern University, Davies, Florida, USA*

²*Florida Atlantic University, USA*

Abstract

For the entire period of recorded 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. In the last few years, researchers have applied computer programs called Neural Networks or Artificial Neural Networks to a variety of uses ranging from medical to financial. The purpose of the study was to demonstrate that Neural Networks can be successfully applied to flood forecasting.

The river system chosen for the research was the Big Thompson River, located in North-central Colorado, United States of America. The Big Thompson River is a snow melt controlled river that runs through a steep, narrow canyon. In 1976, the canyon was the site of a devastating flood that killed 145 people and resulted in millions of dollars of damage.

Using publicly available climatic and stream flow data and a Ward Systems Neural Network, the study resulted in prediction accuracy of greater than 97% 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: flood forecasting, neural networks, hydrologic modelling, rainfall/runoff, hydrology, modelling, artificial neural networks.



1 Introduction

One of the major problems in flood disaster response is that floodplain data are out of date almost as soon as the surveyors have put away their transits.

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

Even current models that rely on linear regression require extensive data cleaning, which is time and data intensive. A new model must be created every time there is a change in the river basin. The process is time, labor, and, data intensive; and, as a result, it is extremely costly. What is needed is a method or model that will do all of the calculations quickly, accurately, using data that requires minimal cleaning, and at a minimal cost. The new model should also be self-updating to take into account all of the changes occurring in the river basin.

With a NN program, a watershed and its associated floodplains can be updated constantly using historical data and real-time data collection from existing and future rain gauges, flow meters, and depth gauges. The constant updating will result in floodplain maps that are current and accurate at all times.

Another problem with published floodplains is that they depict only the 100-year flood. This flood has a 1% probability of happening in any given year. While this is useful for general purposes, it may not be satisfactory for a business or a community that is planning to build a medical facility for non-ambulatory patients. For a facility of this nature, a flood probability of 0.1% may not be acceptable. The opposite situation is true for the planning of a green belt, golf course, or athletic fields. In this situation, a flood probability of 10% may be perfectly acceptable.

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

2 Literature

The term NN is used in this dissertation to represent both the NN and ANN programs.

Muller and Reinhardt [1] wrote one of the earliest books on NNs. The document provided basic explanations and focus on NN modeling. Hertz et al [2] presented an analysis of the theoretical aspects of NNs.

In recent years, a great deal of work has been done in applying NNs to water resources research.

Hjelmfelt and Wang [3] used NNs to unit hydrograph estimation. The authors concluded that there was a basis, in hydrologic fundamentals, for the use of NNs to predict the rainfall-runoff relationship.

Huffman [4] presented a paper that suggested that NNs could be applied to creating floodplains that could be constantly updated without relying on the costly and time consuming existing modeling techniques.



Wei et al [5] proposed using NNs to solve the poorly structured problems of flood predictions.

Rajurkar et al [6] tested a NN on seven river basins. They found that this approach produced reasonably satisfactory results from a variety of river basins from different geographical locations...

Kerh and Lee [7] describe their attempt at forecasting flood discharge at an unmeasured station using upstream information as an input. They discovered that the NN was superior to the Muskingum method.

Filho and dos Santos [8] applied NNs to modeling stream flow in a densely urbanized watershed.

Sahoo and Ray [9] described their application of a feed-forward back propagation and radial basis NN to forecast stream flow on a Hawaii stream prone to flash flooding.

Late in this study, a paper by Hsu et al [10] was discovered demonstrating that results were dramatically improved by adding the previous day's stream flow or stage level input with the other data. This technique was applied in this study. This application resulted in a dramatic improvement of the predictive capability of the model.

3 Methodology

Current methods of stream-flow forecasting are based on in-depth studies of the river basin including (a) geologic studies, (b) topographic studies, (c) ground cover, (d) forestation, and (e) hydrologic analysis. All of these are time and capital intensive.

Several independent variables were considered, and two test bed data sets are used, the Drake and Loveland data sets.

The Drake measuring station is described as, "USGS 06738000 Big Thompson R at mouth of canyon, NR Drake, CO." USGS, [11]. Its location is: Latitude 40°25'18", Longitude 105°13'34" NAD27, Larimer County, Colorado, Hydrologic Unit 10190006. The Drake measuring station has a drainage area of 305 square miles and the Datum of gauge is 5,305.47 feet above sea level.

The Loveland measuring station is described as USGS06741510 Big Thompson River at Loveland, CO. USGS [12]. Its location is Latitude 40°22'43", Longitude 105°03'38" NAD27, Larimer County, Colorado, Hydrologic Unit 10190006. Its drainage area is 535 square miles and is located 4,906.00 feet above sea level. The records for both sites are maintained by the USGS Colorado Water Science Center USGS, [11, 12]. The following data was used in this model:

Tmax is the maximum measured temperature at the gauging site.

Tmin is the lowest measured temperature at the gauging site.

Tobs is the current temperature at the gauging site.

Tmean is the average temperature during the 24-hour measuring period

Cdd are the Cooling Degree Days, an index of relative coldness.

Hdd are the Heating Degree Days, an index of relative warmth.

Prpc is the measured rainfall during the 24-hour measuring period.



Snow1 is the measured snowfall during the 24-hour measuring period.

Snwd is the measured depth of the snow at the measuring site.

The output variable is the predicted flow level.

This is the actual data collected by the meteorological stations. The samples for each site are more than 3000 data sets which are more than enough to (a) run a NN model, (b) to test it, and (c) to validate it. For the same data, a linear regression model using SPSS was run. The same variables dependent and independent were considered (Mazouz [13]).

The Ward Systems product, selected for the research, is the NeuralShell Predictor, Rel. 2.0, Copyright 2000. The following description was taken directly from the Ward Systems website, www.wardsystems.com (Ward Systems Group [14]).

The methods of statistical validation to be used in this paper are as follows: R-Squared, Average Error, and Percent in Range.

4 Analysis and presentation of findings

The following is a topographic map of the Big Thompson canyon. It is a narrow, relatively steep canyon.

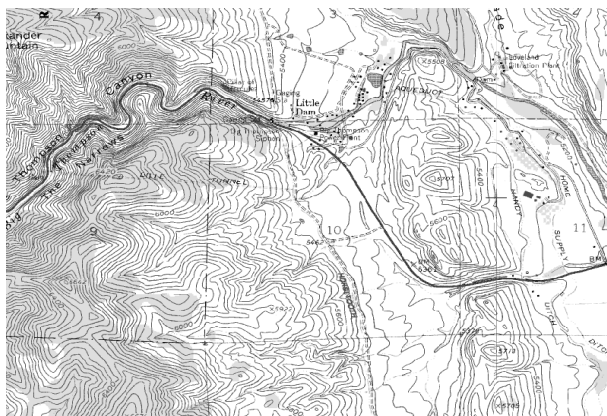


Figure 1: Topography of the Big Thompson Canyon (USGS [15]).

The historical measurements of (a) precipitation, (b) snowmelt, (c) temperature, and (d) stream discharge are available for the Big Thompson Watershed as they are usually available for most watersheds throughout the world. This is in contrast to data on (a) soil characteristics, (b) initial soil moisture, (c) land use, (d) infiltration, and (e) groundwater characteristics that are usually scarce and limited.

For this study, 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 description and locations of the stations are as follows.

<u>Coopid.</u>	<u>Station Name</u>	<u>Ctry.</u>	<u>State</u>	<u>County</u>	<u>Climate Div.</u>	<u>Lat./Long.</u>	<u>Elevation</u>
051060	Buckhorn Mtn 1E	U.S.	CO	Larimer	04	40:37/ -105:18	2255.5
052759	Estes Park	U.S.	CO	Larimer	04	40:23/-105:29	2279.9
052761	Estes Park 1 SSE	U.S.	CO	Larimer	04	40:22/-105:31	2372.9
054135	Hourglass Res.	U.S.	CO	Larimer	04	40:35/-105:38	2901.7
055236	Loveland 2N	U.S.	CO	Larimer	04	40:24 /-105:07	1536.2
058839	Waterdale	U.S.	CO	Larimer	04	40:26/ -105:13	1594.1

NCDC [14]

The period of time for the historical data selected was from July 4, 1990, through May 7, 1998, a total of seven years, ten months and three days.

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.

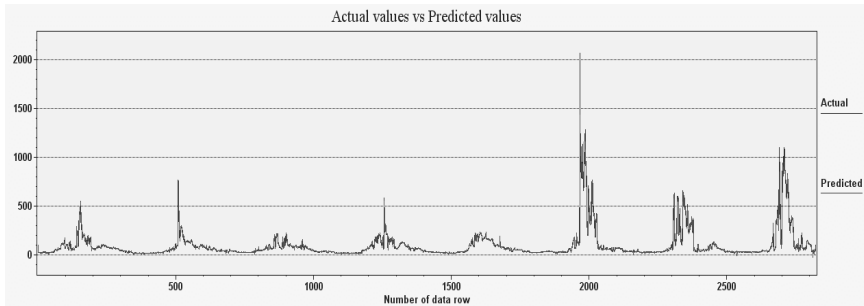


Figure 2: Drake, final model, actual versus predicted.

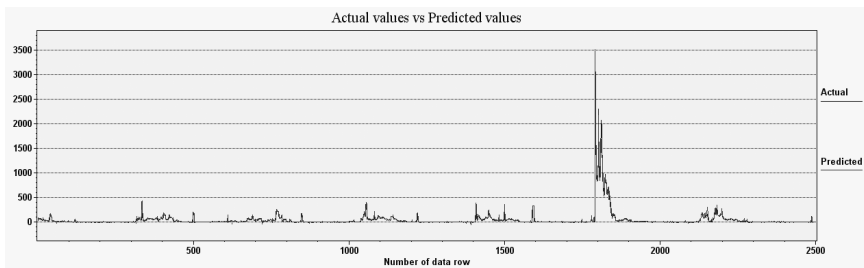


Figure 3: Loveland, final model, actual versus predicted.

The R-Squared results for the Drake and Loveland measuring stations were 0.9091 and 0.9671.

The Average Error for the Drake and Loveland measuring stations are 15.7 cfm and 11.56 cfm.

The Correlation values for both the Drake and the Loveland measuring station for this model are very good at 0.9534 and 0.9834.



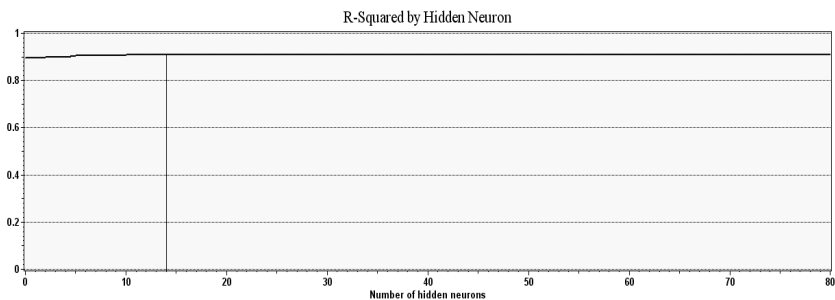


Figure 4: Drake, final model, R-squared.

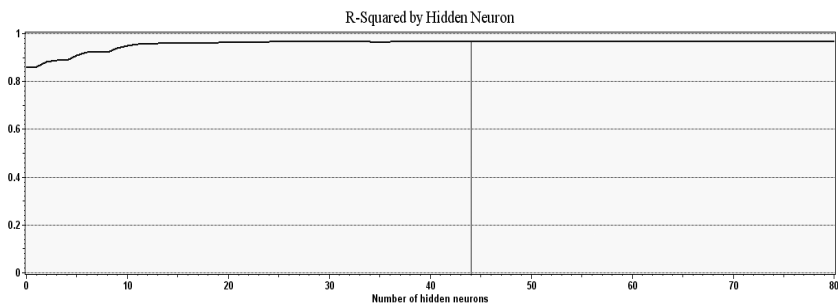


Figure 5: Loveland, final model, R-squared.

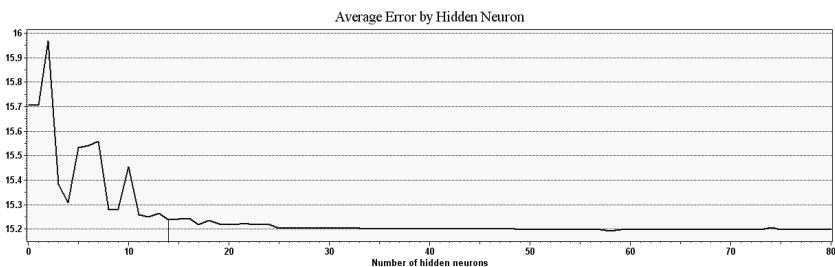


Figure 6: Drake, final model, average error.

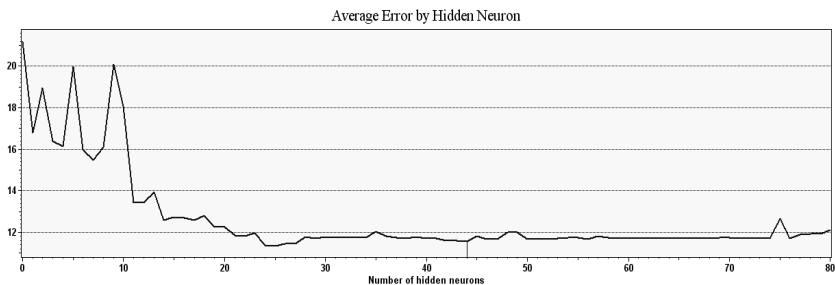


Figure 7: Loveland, final model, average error.



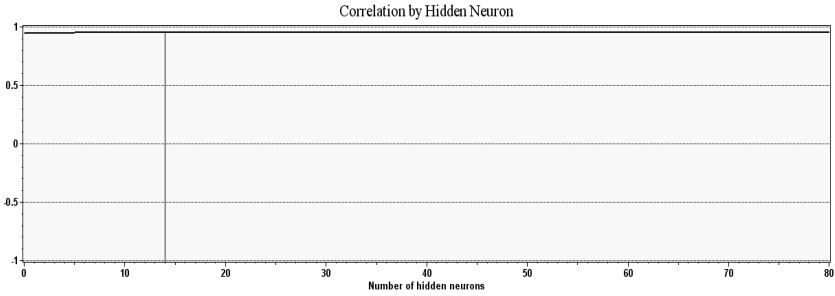


Figure 8: Drake, final model, correlation.

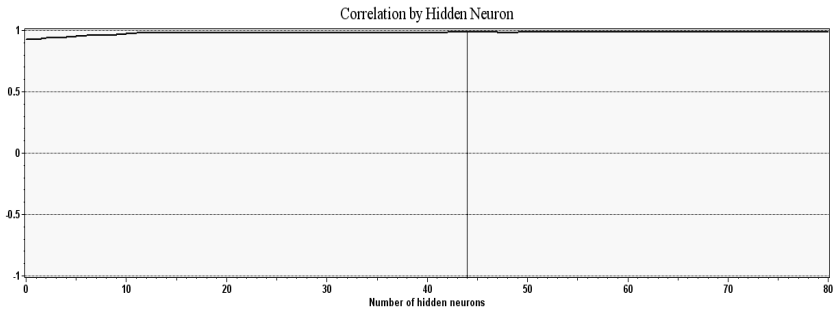


Figure 9: Loveland, final model, correlation.

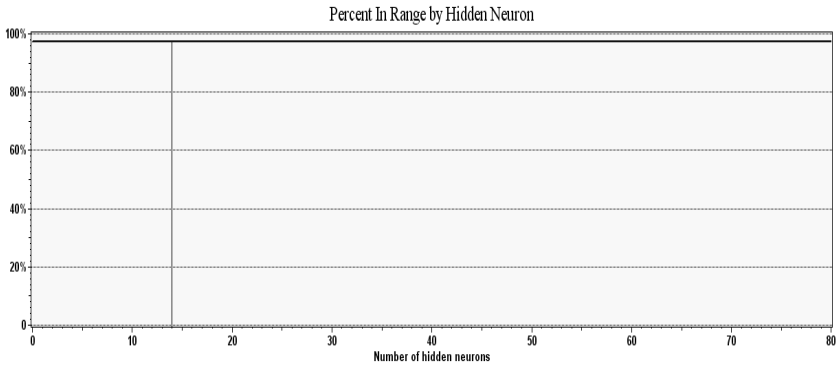


Figure 10: Drake, final model, percent in range.

The Drake and Loveland measuring stations' Percent in Range ended the run at values of 98.1 and 97.3.

The following Multi linear regression models were created and provided by Dr. Kadar Mazouz of Florida Atlantic University (Mazouz [13]).



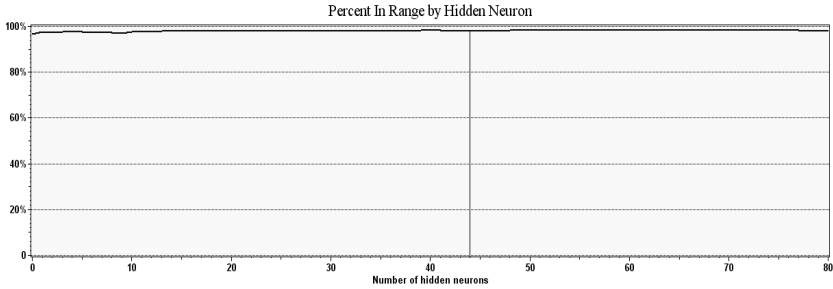


Figure 11: Loveland, final model, percent in range.

Table 1: Drake, model summary (i).

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.916(a)	.838	.838	29.0311
2	.918(b)	.842	.842	28.7095
3	.918(c)	.843	.843	28.6109
4	.919(d)	.845	.845	28.4290
5	.920(e)	.846	.845	28.3747
6	.920(f)	.847	.846	28.3096
7	.921(g)	.849	.848	28.1527

- (a) Predictors: (Constant), Prewflow.
- (b) Predictors: (Constant), Prewflow, Tobs3.
- (c) Predictors: (Constant), Prewflow, Tobs3, Prcp2.
- (d) Predictors: (Constant), Prewflow, Tobs3, Prcp2, Prcp4.
- (e) Predictors: (Constant), Prewflow, Tobs3, Prcp2, Prcp4, Tmax3.
- (f) Predictors: (Constant), Prewflow, Tobs3, Prcp2, Prcp4, Tmax3, Prcp5.
- (g) Predictors: (Constant), Prewflow, Tobs3, Prcp2, Prcp4, Tmax3, Prcp5, Snow5.
- (h) Dependent Variable: OUTPUT.

A stepwise multi-linear regression model was generated for both data sets, Drake and Loveland. Being a multiphase process, it stopped after the seventh model. It gave an R-square of 0.0849, which is less than the NN Model generated for the Drake Data sets.

Actually, the independent variables that contributed the most to the variability of the output are Prewflow, TOBS3, Prcp2, Prcp4, Prcp5, and Snow5. Their corresponding coefficients in the model are:

On the other hand, as for Loveland, the stepwise Multi-linear regression model for Drake was generated in eight iterations. It gave an R-square of 0.80, which is less than the R-square generated for the Loveland data using NNs.

The independent variables that contributed the most to the variability of the output are Prewflow, OFestes, Prcp3, PrcpA, Tmin3, Snow A, Tobs A. Their corresponding coefficients in the model are:

Table 2: Loveland, summary (i).

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.881(a)	0.777	0.776	22.06113
2	0.891(b)	0.793	0.793	21.23824
3	0.892(c)	0.796	0.796	21.09537
4	0.893(d)	0.798	0.797	21.01388
5	0.894(e)	0.799	0.798	20.95246
6	0.895(f)	0.801	0.800	20.85033
7	0.896(g)	0.803	0.801	20.80284
8	0.896(h)	0.803	0.802	20.76851

(a) Predictors: (Constant), Preflow.

(b) Predictors: (Constant), Preflow, OFEstes.

(c) Predictors: (Constant), Preflow, OFEstes, Prcp3.

(d) Predictors: (Constant), Preflow, OFEstes, Prcp3, Prcp_A.

(e) Predictors: (Constant), Preflow, OFEstes, Prcp3, Prcp_A, Tmin3.

(f) Predictors: (Constant), Preflow, OFEstes, Prcp3, Prcp_A, Tmin3, Tmax1.

(g) Predictors: (Constant), Preflow, OFEstes, Prcp3, Prcp_A, Tmin3, Tmax1, Snow_A.

(h) Predictors: (Constant), Preflow, OFEstes, Prcp3, Prcp_A, Tmin3, Tmax1, Snow_A, Tobs_A.

(i) Dependent Variable: LvdFlow

5 Summary and conclusions

In this study, a daily rainfall-runoff model for two flow-measuring stations, Drake and Loveland, on the Big Thompson River in Colorado, was developed using the NeuralShell Predictor. 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 Mazouz [13].

For the Big Thompson River, the NN provides better than 97 % predictions within a plus or minus 100 cfm range. When comparing the results of the NN to those of the linear multiple regression analysis, it is apparent that the NN provides a clearly superior predictive capability.

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.

This model was developed and tested for the Big Thompson River. The Big Thompson River is mainly driven by snowmelt with occasional heavy rains adding complexity. The NN modeling should be tested in other watersheds with different size drainage areas, watershed characteristics, and climatic conditions.



References

- [1] Muller B. and Reinhardt, J., (1990) *Neural Networks, an Introduction*. Springer-Verlag: Berlin and New York, 1990.
- [2] Hertz, J., Krogh, A. and Palmer, R., (1991). *Introduction to the Theory of Neural Networks*. Addison-Wesley Publishing Company, 1991.
- [3] Hjelmfelt, T. A. Jr., and Wang M. *Proceedings of the Symposium on Engineering Hydrology, ASCE, San Francisco, (1993) (unpublished)*.
- [4] Huffman, W. S., (2001) *Geographic Information Systems, Expert Systems and Neural Networks: Disaster planning, mitigation, and recovery*. River Basin Management, pp. 311-321, 2001.
- [5] Wei, Y., Xu, Y., Fan et al., (2002), Artificial neural network based predictive method for flood disaster. *Computers & Industrial Engineering*, 42(204), pp 383-390, 2002.
- [6] Rajurkar, M. P., Kothyari, U. C., and Chaube, U. C., (2004), Modelling of the daily rainfall-runoff relationship with artificial neural network. *Journal of Hydrology*, 285(1-4), pp. 96-113, 2004.
- [7] Kehr, T., and Lee, C. S., (2006), Neural networks forecasting of flood discharge at an unmeasured station using river upstream information. *Advances in Engineering Software*, Article in Press, Corrected Proof, 2006.
- [8] Filho, A. J. P., and dos Santos, C. C., (2006), Modelling a densely urbanized watershed with an artificial neural network, weather radar and telemetric data. *Journal of Hydrology*, 317(1-2), pp. 31-48, 2006.
- [9] Sahoo, G. A., and Ray, C., (2006), Flow forecasting for a Hawaii stream using rating curves and neural networks. *Journal of Hydrology*, 317(1-2), pp. 62-80, 2006.
- [10] Hsu, K. L., Gupta, V., and Sorooshian, S., (1995), Artificial neural network modelling of the rainfall-runoff process. *Water Resources Research*, 31, pp. 2517-2530, 1995.
- [11] USGS (2006 a). USGS 06738000 Big Thompson R at mouth of Canyon, NR Drake, CO. Retrieved November 14, 2006.
- [12] USGS (2006 b) USGS 06741510 Big Thompson River at Loveland, CO. Retrieved November 14, 2006.
- [13] Mazouz, A. K., (2006), Multiple Regression Analysis of the Big Thompson River, 2006 (unpublished).
- [14] Ward Systems Group. I., NeuroShell Predictor, Frederick, MD: Ward Systems Group, 2000.
- [15] USGS & MAAC Inc, (2006), Topographic map of the Big Thompson Canyon, 2006. Retrieved November 14, 2006 from www.topozone.com/map.asp?lat=40.41559&lon=-105.223&size=l&u=4&datum=nad27&layer=DRG

