APPLICATION OF ARTIFICIAL NEURAL NETWORK MODEL TO FORECAST RUNOFF FOR WAIKATO RIVER CATCHMENT

MR. ZHIMIN ZHENG[1], DR. BABAR MAHMOOD[2]
[1] Student, [2] Staff at Unitec Institute of Technology, Auckland, New Zealand

AN OVERVIEW

- Acknowledgement from the first author
- Introduction/Background – What and Why?
- Aim of the project
- Methodology (*a brief introduction to ANN is included here*)
- Results & Discussion
- Key findings of the study

ACKNOWLEDGEMENTS

- I would like to extend my thanks to my mentor, Dr Babar Mahmood, for his support and guidance from beginning to the completion of this project. I appreciate the technical knowledge and materials he provided, which was very helpful towards the completion of this project on time.
- I would also like to extend my appreciation to the Waikato Regional Council for providing me with data, so that I could complete this project.
- Acknowledgement is also due to Dennis Sewpershad who is Autodesk ATC Manager at the Faculty of Technology and Build Environment, at Unitec. He helped to facilitate me to work on Matlab on my laptop.

WHAT & WHY

What?

- Timely prediction of surface run-off is very important as we all know.
- The rainfall-runoff relationship is considered to be one of the complex hydrological processes to be modelled.

Why? –

  - Because of the involvement of a number of variables in the modelling process
  - Enormous spatial and temporal variability of watershed/catchment characteristics

- Empirical models are not universal – means not used and understood by all.
- Numerical models have been used by engineers/planners – but are bit data hungry.
- Therefore, there was need for user friendly and more robust models that requires less data to run.
- Artificial Intelligence (AI) - techniques (e.g. ANN - Artificial Neural Network & FIS - Fuzzy Interference System) have been used as an alternative modelling tools in water resources management studies.
- ANN technique has been identified as a powerful technique that could be used to model systems where the explicit form the relationship between variable(s) is unknown.
- This study is about the use of ANN model to predict runoff using the past rainfall-runoff data for a catchment.
- A brief introduction to the model is provided in the methodology section.

AIM & OBJECTIVES

Similar to other modelling approaches, successful application of ANN is **dependant** on the selection of appropriate input factors - ***To investigate this***, *the study applied three different approaches for the selection of appropriate input vectors to be used for the ANN model*".

Therefore, the specific objectives of the study were to:

1. Learn the fundamentals of the Neural network tool box of Matlab software.
2. Investigate different approaches available to determine the optimum input vectors for the ANN model.
3. Check the sensitivity of ANN model performance to the selection of appropriate input vectors.
4. Identify the best ANN model by comparing measured & predicted flow duration curves.

METHODOLOGY

- **ANN** is a computing paradigm and inspired by the working of human brain nervous system.
- Neurons connections is called <u>**synapse**</u>.
- Theses neuron produce proper response by releasing chemicals, which can cause a synapse to conduct an electric current.
- Neurons are considered the main building block of ANN.
- **FFNN** (one of the types of ANN)
  - Three layers
  - Each neuron is connected with neuron in the next layer through connection called <u>**'WEIGHTS'**</u>.
  - X_1, X_2, X_n are the input parameters
  - $\mathbf{W^H}$ $\mathbf{_{12}}$ - refers to the connection weight between the 1st neuron in hidden layer and 2nd neuron in the input layer.
  - $\mathbf{W^O}$ $\mathbf{_{12}}$
  - Each neuron in each layer processes its information by a function called '<u>**neuron transfer function**</u>'.
  - No direct connect between neuron in the input layer to neuron in the output layer.

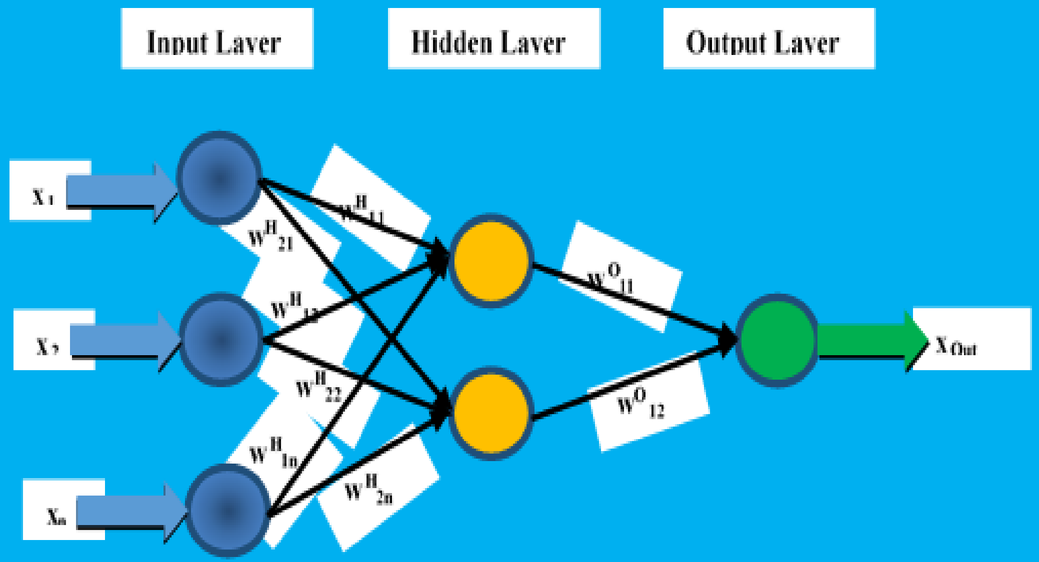

Figure 1. Feed-Forward Neural Network (FFNN).

METHODOLOGY…….CONTINUED

- **Study area**: Waikato catchment area of 14500 km^2
  - 425 km in length, longest river of NZ
  - Originates from Lake Taupo and discharges to Tasman sea
- **Input Data**:
  - Daily rainfall and discharge data was obtained from Environment Waikato
  - 10 years of data starting from May 2002 was used in this study
  - Data from 16 rainfall stations was used to calculate the average rainfall depth for the catchment
- Neural network type used was **FFNN** (being the most common network used in hydrological studies)

METHODOLOGY……CONTINUED

Selection of Transfer Function:

- An Important parameters of ANN, which controls the generation of output in a neuron.
- A default hyperbolic function was used in this study. This transfer function is continuous and differentiable and monotonically increasing.

Selection of Hidden Layer Neurons:

- Performance of ANN is very sensitive to the selection of a number of neurons in the hidden layer in order to obtain the reliable results
- Appropriate number of hidden neurons was completed by trial and error method as reported by Nourani et al., 2011; Shamseldin, 1997; Adamowski & Sun, 2010)
- This trial & error procedure involves training of the network and evaluating its performance.
- Therefore, the selected range was 5 to 40 to find the best results.

Selection of Initial Weights and Stopping Criteria:

- The initial weights were between 1 and -1 in the ANN model.

Selection of Learning Algorithm:

- The LMA (Levenberg-Marquardt Algorithm) available in the neural network tool box was used because it was quicker and more reliable (Beale, et al., 2015)

Selection of Input Vectors is covered in the results & discussion section

Development of Flow Duration Curves (FDC):

- Testing of the developed ANN model was undertaken by preparing the low, medium and high flow duration curves using measured data – [using (Klingeman, 2005) method].
- Also, statistical measures (i.e. correlation coefficients, root mean square error -RMSE, and Nash-Sutcliffe efficiency - NSE) were used to assess the performance of the developed ANN model.

RESULTS & DISCUSSION

Selection of Input Vectors: The performance of rainfall-runoff data driven models (including FFNN type of ANN) is very much dependant on selection of appropriate input data set..

- Three approaches were found in the literature:
  - **Approach 1** – Sequential time series - *Input vectors were selected starting with 1-day lagged time series and then successively adding one more lagged time series into input vectors, and it continues up to a specific lag time*.
  - **Approach 2** – Pruned time series, - *Input vectors were selected based on a cross-correlation between obs. rainfall & runoff data*
  - **Approach 3** – Non-sequential time series (***refer*** to Section 3.1 for more details) - *Input vectors were selected on the basis of cross and auto correlation analysis of the data.*

As an example:

- The ten input vectors i.e. I_1, I_2, …. I_9 and I_{10} are presented below to show an example of the input vector selection:
- **I_1 = r(t)**
- **I_2 = r(t), r(t-1) ………**
- **I_{10} = r(t), r(t-1), r(t-2), r (t-3), r(t-4), r (t-5), r(t-6), r (t-7), r(t-8), r (t-9)**
- The first input vector contains only one day antecedent rainfall data in the input vector.
- The second vector contains lagged 1 and lagged 2-day rainfall, the third input vector contains lagged 1, lagged 2 and lagged 3-day rainfall and so on to forecast the runoff.
- The first input vector I_1 contains only one variable which is the lagged 1-day rainfall data. The input vector I_2 contains two variables of lagged 1 and lagged 2-day rainfall data and so on.

RESULTS & DISCUSSION …..**APPROACH 1**

Least performing model

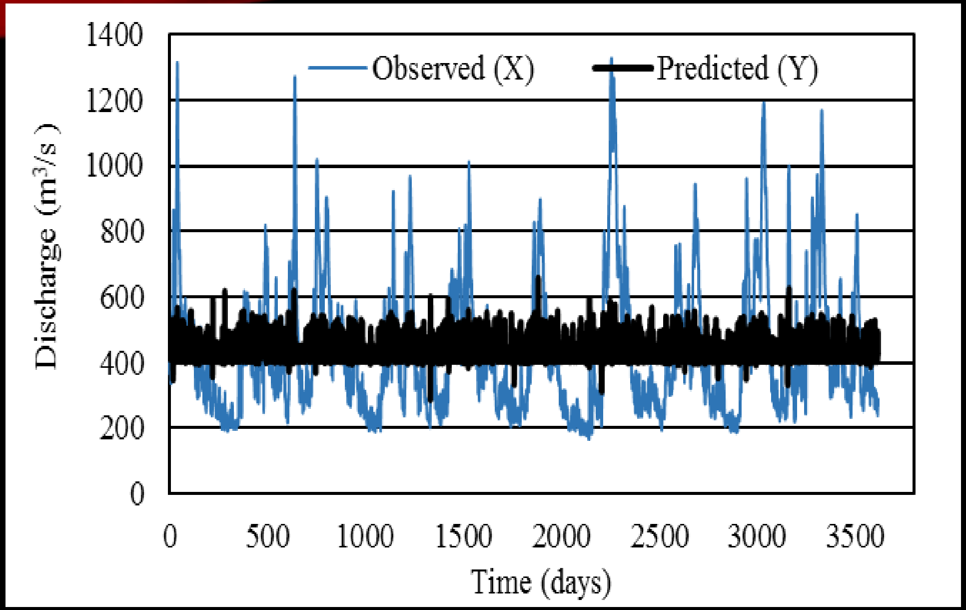

Figure 2. The performance of the ANN model (i.e. observed vs predicted flow rates) for input vector (I_1) with sequential time series approach.

Best performing model

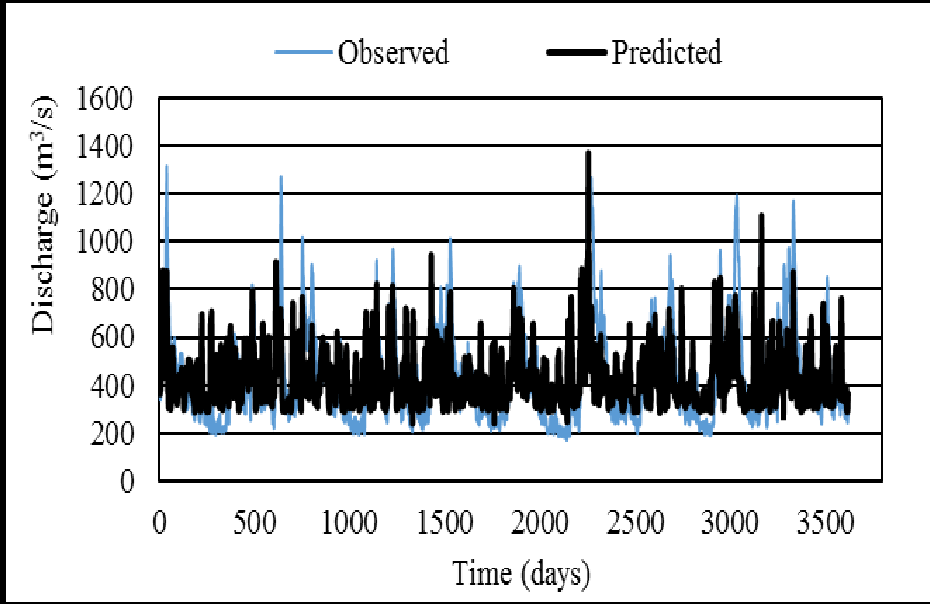

Figure 3. The performance of ANN model (i.e. observed vs predicted) for I_{10} with sequential time series approach.

RESULTS & DISCUSSION …..**APPROACH 2**

Least performing model

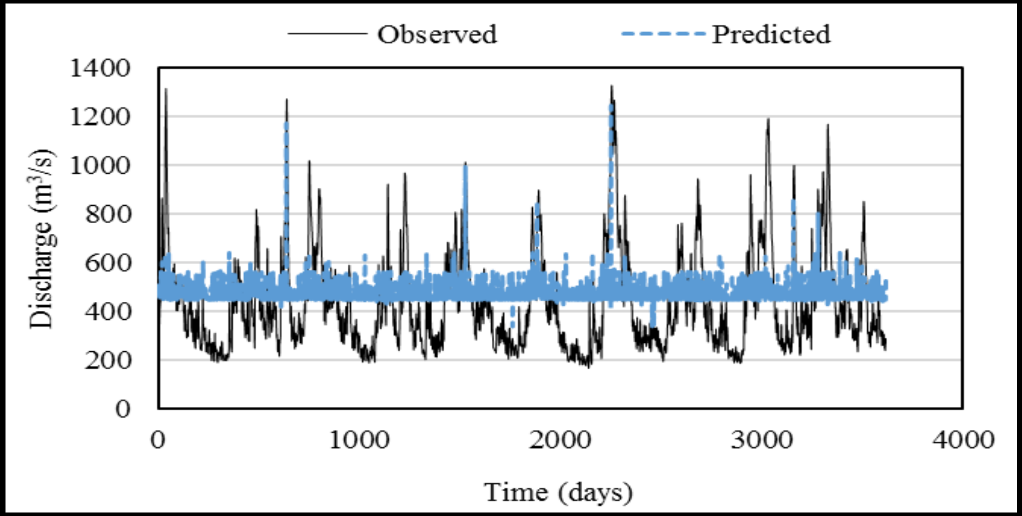

Figure 4: The performance of ANN model (i.e. observed vs predicted daily flow rates) for I_1 with pruned time series approach.

Best performing model

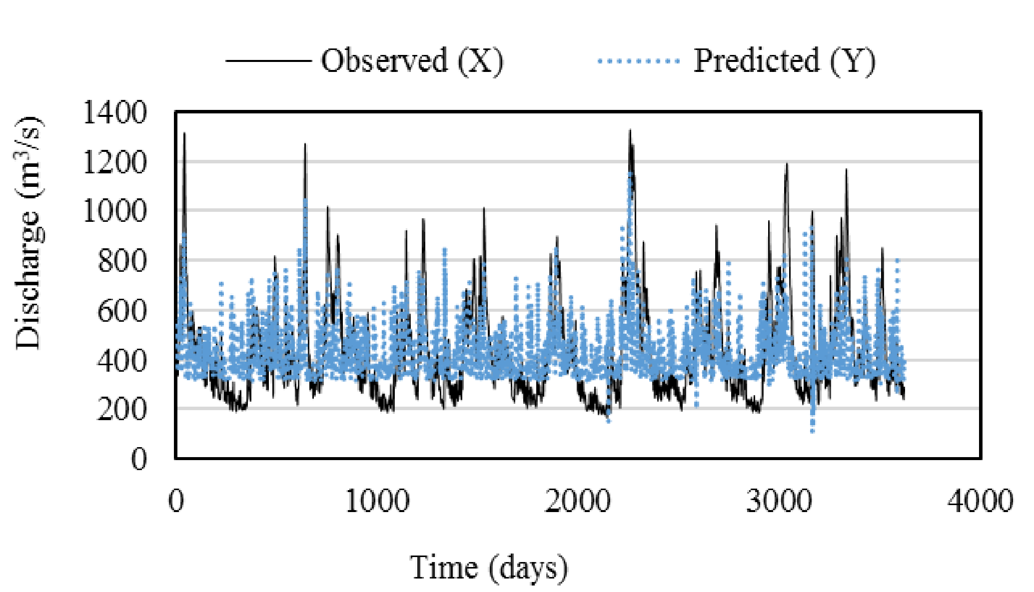

Figure 5: The performance of ANN model (i.e. observed vs predicted daily flows) for I_8 with pruned time series approach.

RESULTS & DISCUSSION …..**APPROACH 3**

Least performing model

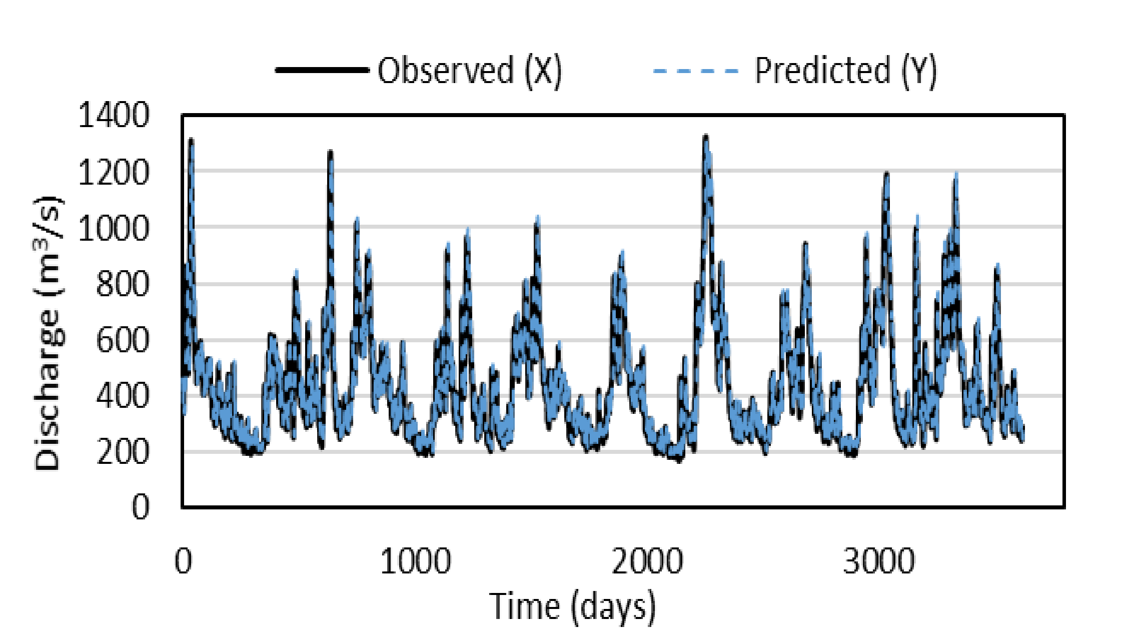

Figure 6: The performance of the ANN model for I_1 with non-sequential approach.

Best performing model

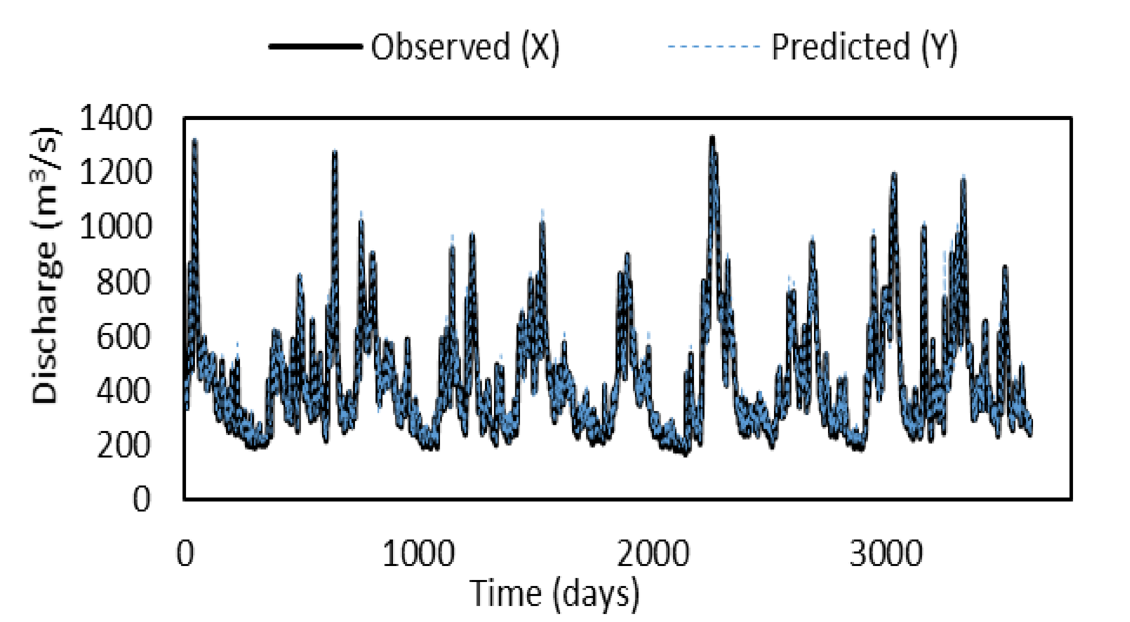

Figure 7: The performance of the ANN model developed using I_5 with non-sequential approach.

- Performance Comparison of ANN Models

Table 1: Performance parameters (i.e. R^2, RMSE and NSE) of ANN models (with different input vectors) for three approaches (i.e. sequential, pruned, and non-sequential time series, respectively).

		Training			Validation			Testing		
Approach	Input	R^2	RMSE	NSE	R^2	RMSE	NSE	R^2	RMSE	NSE
	Vectors	(%)	(m^3/s)	(%)	(%)	(m^3/s)	(%)	(%)	(m^3/s)	(%)
1	I_1	5	195	5	10	220	8	6	195	5
	I_{10}	36	161	36	39	179	39	35	162	35
2	I_1	11	197	9	10	215	7	8	194	8
	I_8	29	169	29	27	199	26	22	180	20
3	I_1	97	35	97	97	38	97	95	43	95
	I_2	97	35	97	97	37	97	97	37	97
	I_3	98	31	98	98	32	98	95	43	95
	I_4	98	31	98	98	32	98	97	36	97
	I_5	98	31	98	98	32	98	97	37	97

RESULTS AND DISCUSSION

- Measured and Predicted Flow Duration Curves for the Three Approaches

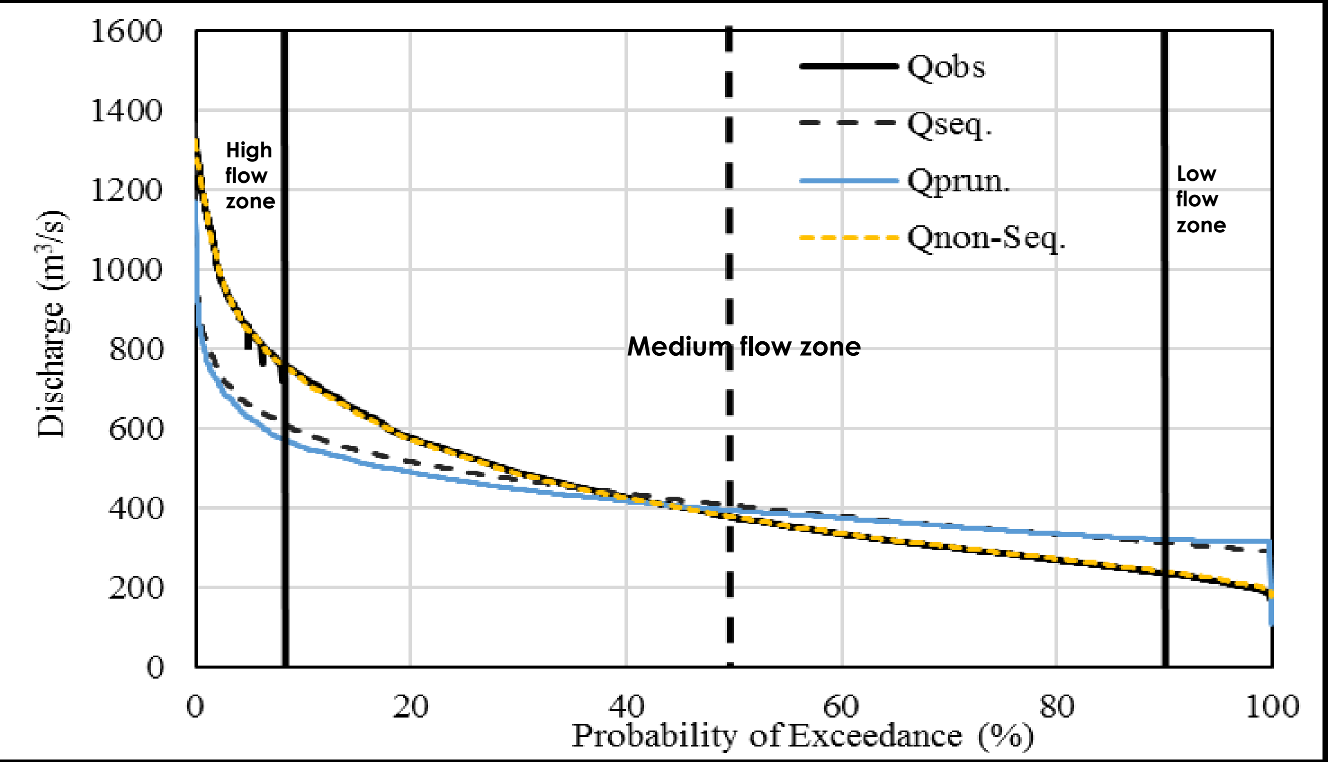

Figure 8: Observed and predicted flow duration curves for three approaches (i.e. sequential, pruned and non-sequence time series).

ANN model developed with Non-sequential time series approach

was

considered to be the BEST MODEL for this project

SUMMARY & CONCLUSIONS

1. Generally speaking, the ANN models developed using three approaches (i.e. *sequential*, *pruned* and *non-sequential* time series) were able to predict runoff using the antecedent rainfall data.

2. However, it was found that the ANN model performance was very sensitive to the selection of appropriate input vector. As the input vector values increases the model performance increases.

3. Input vectors selected on the basis of cross correlation (i.e. statistical dependence of two variables such as rainfall and runoff in this case) and auto correlation (i.e. cross relation of discharge with its lagged values) yielded the best results.

4. It is clear on the basis of the above analysis that the ANN model developed with I_{10} in *sequential time series* approach yielded the best results among the ten input vectors tested. Likewise, I_8 and I_5 produced the best results with the *pruned time series* approach and the *non-sequential* time series approach, respectively.

5. The ANN model developed using approach 3 (i.e. *non-sequential* time series) performed well, and gave, on average, 96% value for both R^2 and NSE during the validation and testing processes.

RECOMMENDATIONS

- Further, ANN models developed in this study were not able to provide an estimation of other important hydrological processes of the catchment like evapotranspiration, infiltration rate of the soil, sub-surface flow and groundwater flow etc.
- It is, therefore, recommended to add some other important hydrological parameters like evaporation, slope and area of the catchment in the input vector. This may improve the predictions of the ANN model for *sequential* and *pruned* time series approaches.