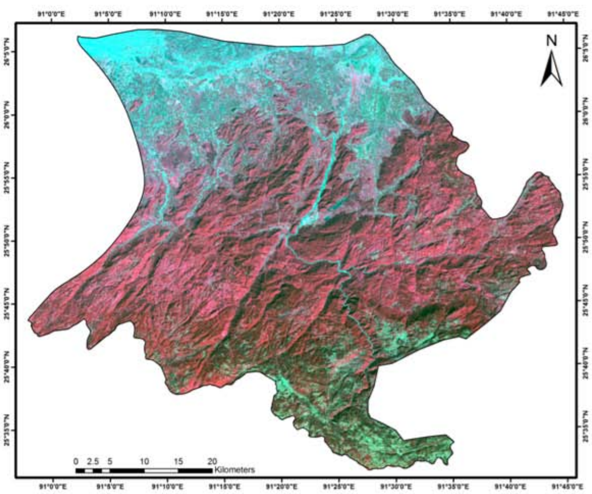
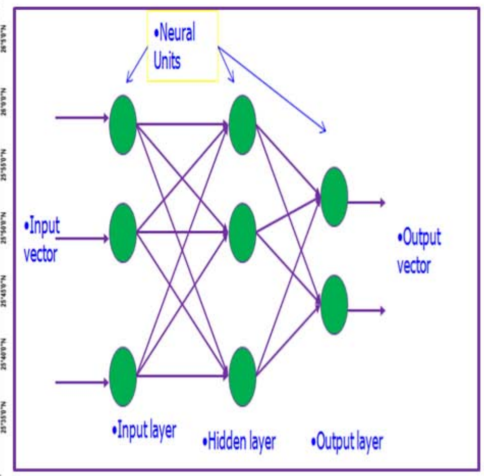
SHORT TERM FLOOD FORECASTING USING BOOTSTRAP BASED ARTIFICIAL NEURAL NETWORKS WITHIN KULSI RIVER BASIN (ASSAM / MEGHALAYA)





आपो हिष्ठा मयो भुवः

NATIONAL INSTITUTE OF HYDROLOGY
Centre for Flood Management Studies
Dispur, Guwahati – 781 006
2015-17

PREFACE

River flow forecasting deals with the estimation of future stages or flows at a single or multiple sites of a river system. Daily river flow forecasts are essential for water resources planning and management including potential water supply for domestic needs, irrigation scheduling, hydropower generation and regulating flows through reservoirs and barrages, whereas hourly water level forecasts are essential for issuing flood warning to mitigate natural disaster by undertaking appropriate evacuation and rehabilitation plans. The necessity of accurate and reliable river flow forecasts is increasingly being felt with increasing demands on water resources due to economic development and demographic expansion.

Integrated Water Resources Management (IWRM) is a process which promotes the coordinated development and management of water, land and related resources, in order to maximize the resultant economic and social welfare in an equitable manner without compromising the sustainability of vital ecosystems. National Institute of Hydrology has undertaken a project under XIIth Plan at each of its Regional Centres.

The Kulsi River Basin, a part of the Brahmaputra sub-basin is situated on the south bank of the mighty River Brahmaputra was identified as Pilot Basin for Centre for Flood Management Studies, Guwahati. This sub-basin spreads in the Kamrup District of Assam as well as west Khasi hills and Ribhoi district of Meghalaya. The river Kulsi drains out a total area of 2806 sq. km. within the Kamrup District of Assam as well as west Khasi hills and RiBhoi district of Meghalaya.

The present report has been prepared to collate and review the progress of Bootstrap based Artificial Neural Networks (BANNs) in order to examine the feasibility of application of Neural Networks in conjunction with Geographic Information Systems (GIS) for flood forecasting.

The report has been prepared by Dr. Sanjay Kumar Sharma, Sc. ‘B’, Er. Gulshan Tirkey, Sc ‘B’ and Dr. C. K. Jain, Sc ‘G’ & Head, Centre for Flood Management Studies, Guwahati, under the work programme for 2015-17.

(S.K. Jain)
Director

CONTENTS

LIST OF FIGURES

LIST OF TABLES

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 (ANNs) 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 ANNs are recognized as being universal approximators and are capable of extracting the underlying relationship between any input, or stimulus, and its subsequent output, or response. The ANNs compose of layers with many nodes in each layer. Input data are fed into the first layer called the input layer, while the outputs are taken from the last layer, called the output layer, and the layers in between are hidden layers. To employ the ANNs, the network architecture must be chosen beforehand.

The bootstrap method is a computational procedure that uses intensive resampling with replacement to reduce uncertainties. The aim of re-sampling is to mimic the random component of a process and to reduce variance through averaging over numerous different partitions of the data. It is commonly used to estimate confidence intervals, but it can also be used to estimate bias and variance of an estimator or calibrate hypothesis tests. Bootstrapping generates multiple versions of predictor to obtain an aggregated predictor. The aggregation averages over the versions when predicting a numerical outcome by making bootstrap replicates of the learning sets and using these as new learning sets

The present study makes an attempt to improve the credibility of the data-driven models among researchers and practitioners. Earlier studies on NN modeling have not given much emphasis on ensemble modeling and uncertainty assessment in hydrologic forecasting for different lead times. The combined strength of bootstrap method with ANN for hydrologic forecasting has not been explored. Also, the uncertainty assessment for longer lead times has not been carried out.

Keeping the above in view, the present research is undertaken with the following objectives:

1. To explore the potential of flood forecasting using ANN in Kulsi River Basin &
2. To quantify the uncertainty in flood forecasts using bootstrap technique.

1.0 INTRODUCTION

Flood disasters continue to occur in many countries around the world and cause tremendous casualties and properties damage. To mitigate the effects of floods, both structural and non-structural measures can be employed, such as dykes, channelization, flood proofing of properties, land-use regulation and flood warning schemes. Sufficient advance warning time may save lives and property by allowing time to effect various structural and other adjustments. Earlier advanced warning can be achieved through advances in mathematical modeling. There are many rainfall-runoff models being developed and employed for flood forecasting which lead to the issue of flood warnings. Artificial intelligence techniques have recently been introduced, such as Artificial Neural Networks (ANN).

The ANNs are recognized as being universal approximators and are capable of extracting the underlying relationship between any input, or stimulus, and its subsequent output, or response. The ANNs compose of layers with many nodes in each layer. Input data are fed into the first layer called the input layer, while the outputs are taken from the last layer, called the output layer, and the layers in between are hidden layers. To employ the ANNs, the network architecture must be chosen beforehand. Many examples of applying ANN-models were given by Hall & Minns (1993), Minns& Hall (1996), Campolo et al (1999), Dawson &Wilby (1998) and Cameron (2002). The recent research on ANNs as hydrological models has offered the opportunity to avoid prior processing, for example the effective rainfall (total rainfall minus losses), since ANNs are not constrained by volume continuity. Nor does a complex physical structure have to be established prior to modeling. The ANNs offer not only a relatively quick and flexible means of modeling (Imrie et al, 2000), but also a way to find relationships between different input samples (Dawson &Wilby, 1998).

Although papers describing interest in the application of ANNs as hydrological models only began to appear in the early 1990s, a considerable body of literature has now accumulated. Since comprehensive reviews have been provided by ASCE Task Committee (2000a,b), Maier and Dandy (2000), Dawson and Wilby (2001) and Minns and Hall (2003), only a brief overview of these techniques will be given here.

There are three broad categories of problems that are involved in the use of ANNs as general modeling tools, and their particular employment for forecasting purposes, namely:

a) The overall architecture of the neural network, and the algorithms required to obtain the values of their weights;

b) The ability of the neural network to extrapolate beyond the range of its training data set; and
c) The selection of the inputs to the ANN that are required to represent the 'pattern' which gives rise to the 'pattern' of outputs.

Although the advantages of different architectures of network in specific applications are acknowledged, this study concentrates on the use of Multi-layer Perceptrons (MLPs) whose weights are obtained with the back-propagation algorithm.

The solutions to the second and third of these problems are fundamentally dependent upon the application of the modeller's domain knowledge, and form the principal focus of this study. Minns and Hall (2003) have noted that the first choice to be made by the ANN modeller is the mode of presentation of the data to the network, i.e. how can the input and output patterns to be defined? One possibility is to take the n successive ordinates of the rainfall hyetograph and feed these into the n input nodes of a network whose m output nodes carry the m successive ordinates of the flow hydrograph (Lange, 1998).

An alternative strategy is the so-called dynamic approach in which the input is a set of concurrent ordinates of (say) the rainfall totals from p raingauges within the catchment and the output is the concurrent rate of outflow. In this mode, each time step defines a pattern, and therefore a single storm event provides as many exemplars as there are runoff ordinates, rather than only a single input-output pairing. This is the approach that has been adopted by the majority of writers on rainfall-runoff modeling using ANNs, but requires a much higher level of hydrological insight into the working of the catchment system. Within the dynamic approach, three types of model can be identified:

The dynamic model:

$$Q(t) = f(r(t), r(t-1), K \quad (1.1)$$

The rainfall-runoff simulation model:

$$Q(t) = f(r(t), r(t-1), K, Q(t-1), Q(t-2), K \quad (1.2)$$

The auto regressive model:

$$Q(t) = f(Q(t-1), Q(t-2), K \quad (1.3)$$

Where Q(t) is the outflow at time interval t and r(t) is the rainfall ordinate at time t

The simplest, naïve dynamic rainfall-runoff model Eq(1.1) would consist of an ANN with inputs from one or more raingauges at time and an output of concurrent flow. A simple scatter plot of input(s) against output is sufficient in this case to indicate that the description of the input pattern is inadequate. Some improvement is obtainable by allowing for the time lag between the occurrence of the flow and the incidence of the causative rainfall. Since the flow at any instant is effectively composed of contributions from different sub-areas whose time of travel to the outlet covers a range of values both the concurrent and antecedent rainfalls can be considered to be contributing to the outflow. Use of a moving window of rainfall at time t and the k previous intervals provides some improvement. The choice of length for the moving window of rainfall can significantly affect the accuracy of the resulting ANN model. If the window is too short, the input data does not contain enough information about the entire rain event that is contributing to the concurrent outflow. Conversely, if the window is too long, the input contains too much information, which may even include historical rainfall events whose effects have long since passed out of the catchment and so are no longer contributing to the concurrent flow.

A further problem with model eq (1.1) is that the simplistic input patterns may easily result in ambiguous results. More specifically, intervals with zero rainfall inputs are encountered in two contrasting situations: immediately prior to the beginning of the storm and the start of the rising limb; and sometime after the end of the storm event when flows are moderately high and in recession. The ANN has no information to discriminate between these two 'no-rainfall' conditions and once more becomes 'confused'. These conditions have to be differentiated by the addition of another input if the ANN is to achieve the correct mapping. The most obvious candidate is a flow ordinate, which is most easily provided by the output at time t to become an input at time t+l and similar use of stage outputs by (Campolo et al, 1999). This approach is described above as model eq (1.2). In effect, the flow (or stage) ordinate is employed as a crude measure of catchment wetness, although other researchers prefer to use a moving average of flow and/ or rainfall as an alternative measure of wetness (Dawson &Wilby, 1998).

Model eq(1.3) is then the logical extension of models eq (1.1) and eq (1.2) into purely 'autoregressive' time series prediction. This model does not make use of any rainfall data at all but uses only antecedent outflow values as input to the ANN to predict the concurrent outflow. This ANN model provides only slightly poorer performance than the results from model eq(1.2), but suffers from an inability to capture the timing of peak flows. At the peak, the network has no other information available that tells it at which level the rising limb should stop until the actual measurements indicate that this is so. A similar phase error occurs at the beginning of the recession limb, caused by the ANN having no knowledge about the cessation of the rainfall until one or two time steps after the actual measured values start to decrease.

In general, the literature involving the application of ANNs to rainfall-runoff modelling confirms the exceptional accuracy of ANN models for short forecasting intervals. Longer forecasting intervals may be obtained by utilizing Q(t+1), Q(t+2),...etc. as outputs during the training of the ANN. Unfortunately, the performance of the ANN decreases quite rapidly with an increasing prediction time horizon (Campolo et al., 1999, Zealand et al., 1999). Another approach is to use a trained ANN with a 'feedback' loop in which the predicted output is used directly as input data for the subsequent time step. Unfortunately, the error accumulation associated with this approach also means that the performance of the ANN deteriorates quite rapidly after only one or two iterations. The more promising approach would appear to involve the use of (partial) recurrent neural networks, which contain feedback loops in both the training and recall modes of the network show significant improvements in long-term predictions using recurrent neural networks, which they rather unfortunately refer to as 'auto regressive' neural networks.

1.1 IWRM

IWRM is a tool for achieving the Millennium Development Goals (MDGs) which recognises the role of water to achieve development goals. The Plan of Implementation adopted at the World Summit on Sustainable Development in Johannesburg in 2002 called for countries to “develop Integrated Water Resources Management and Water Efficiency Plans by 2005”. These “plans” are milestones in recurring and long-term national water strategy processes.

IWRM provides the means of balancing and meeting the needs for use of water resources in such a way as to ensure the equitable and sustainable use of the water resource. It is based on the principle that, in order to maximise the benefits of the water resource and to ensure equitable use of water, you must balance the needs of all the water users (and discharges) in the catchment.

In achieving the above, the following results were aimed at:

- increased awareness on importance of environmental approach and considerations in IWRM
- increased access to relevant IWRM information and tools
- targeted training for key managers and decision makers in the water sector
- development of a Roadmap for the implementation of IWRM concepts in future planning processes
- guide the River Basin Authority to spearhead the implementation of IWRM plans and inclusion of IWRM in decision making

- prospect for implementation of a baseline study for the development of an IWRM Action Plan by local specialized organization
- Documentation on best practices, case studies and guidelines to enhance replication.

The pilot projects are being encouraged and using a participative approach to develop identification of issues and outcomes to formulate and establish consensus on visions, strategies, outputs, activities and external factors for the different components of the project as well as formulating and identifying the various factors influencing the long-term sustainability of the program.

The basis of IWRM is that different uses of water are interdependent. Additional benefits can be derived when different user groups are consulted in the planning and management of water management programs as such users are likely to apply local self-regulation in relation to issues such as water conservation and catchment protection far more effectively than central regulation and surveillance can achieve.

IWRM is an important instrument to address poverty reduction. Good water governance, the objective of IWRM, and the objective of any “IWRM plan”, is to ensure wise water governance which contributes to the economic development, social equity and environmental sustainability of the society. Implementing an IWRM process is a question of getting the “three pillars” right: (1) moving towards an enabling environment of appropriate policies, strategies and legislation for sustainable water resources development and management; (2) putting in place the institutional framework through which to implement the policies, strategies and legislation; and (3) setting up the management instruments required by these institutions to do their job.

National Institute of Hydrology (NIH) is proposing a project on Integrated Water Resources Management (IWRM) under Pilot Basin Study (PBS) at each of its Regional Centres, during the XII plan period.The PBS programme will involve identification of suitable basin in consultation with concerned State Govt. authorities, establishment of advanced instrumentation for data collection and storage, data processing and analysis using state of the art models, preparation of results and findings in collaboration with specialists of other relevant disciplines in a meaningful and usable form for the intended beneficiaries, organisation of capacity building and awareness programmes for the beneficiaries and stake holders and so on.

The Kulsi River Basin, a part of the Brahmaputra sub-basin is situated on the south bank of the mighty River Brahmaputra was identified as Pilot Basin for Centre for Flood Management Studies, Guwahati This sub-basin spreads in the Kamrup District of Assam as well as west Khasi hills and Ribhoi district of Meghalaya. The river Kulsi drains out a total area of 2806 sq. km. within the Kamrup District of Assam as well as west Khasi hills and RiBhoi district of Meghalaya.

The present report entitled “reviews the application of BANN for prediction of hydrometerological parameters and explores the estimation of precipitation at a discharge site using spatial distribution of selected weather stations.

1.2 River Flow Forecasting using ANN

River flow forecasting deals with the estimation of future stages or flows at a single or multiple sites of a river system. Daily river flow forecasts are essential for water resources planning and management including potential water supply for domestic needs, irrigation scheduling, hydropower generation and regulating flows through reservoirs and barrages, whereas hourly water level forecasts are essential for issuing flood warning to mitigate natural disaster by undertaking appropriate evacuation and rehabilitation plans. The necessity of accurate and reliable river flow forecasts is increasingly being felt with increasing demands on water resources due to economic development and demographic expansion.

Classical time series models such as auto regressive integrated moving average (ARIMA) are widely used for hydrological time series forecasting as they are accepted as a standard representation of a stochastic time series (Maier and Dandy, 1997). However, these models are basically linear models which make use of classical statistics to analyse the historical data. They assume that data are stationary and have limited ability to represent non-stationary and nonlinear dynamics, if any, between the input and output variables. However, most hydrological processes exhibit high nonlinearity between the input and output variables, and hence, linear time series models may not always perform well. In the past, owing to the difficulties associated with nonlinear model structure identification and parameter estimation, the usual practice was to assume linearity or piecewise linearity in modelling nonlinear hydrological processes (Hsu et al., 1995).

To overcome the limitations of classical time series models, a wide variety of rainfall-runoff models have been developed and applied for river flow forecasting ranging from complex physically based to simple black box models. Black box models in the form of neural networks (ANNs) have gained momentum in last few decades for river flow forecasting and have been accepted as a good alternative to physically based and conceptual models (ASCE, 2000a,b). The ability of NNs in extracting the complex nonlinear relationship between inputs and outputs without explicitly accounting for the physical processes has increased the number of applications in river flow forecasting. Most importantly NN models need limited inputs such as water level, discharge, rainfall or sometimes only a single input, whereas physically based models require several additional parameters which are difficult to measure because of temporal and spatial variability. This approach has also been criticized for making models overly complex which lead to problems of over parameterisation and equifinality (Beven, 2006) causing large prediction uncertainty.

NN models are computationally fast and efficient, which makes them a very suitable tool for river flow forecasting. Disadvantages related to NN models include interpretation of the NN structure ("black box") and their extrapolation capacity (Minns and Hall, 1996). Recently, researchers have been exploring different pre-processing approaches for inclusion of additional hydrological knowledge as input to NN models to improve the hydrological representation and generalisation (Corzo and Solomatine, 2007). The quantification of the uncertainty associated with the results provided by NN models is essential for their confident and reliable use in practice as operational river flow management strongly depends on the accuracy and reliability of flow forecasts. Uncertainty assessment in real-time river flow forecasts is an important means for increasing the applicability of hydrological forecasts to mitigate the natural disaster. A probability distribution function of a predictand, or an ensemble of possible realisations of a predictand, enables users to make decisions that take the risk explicitly into account (Georgakakos and Krzysztofowicz, 2001).

The reliability of the model estimated discharge is affected by three sources of uncertainties (Bates and Townley, 1988): data uncertainty (quality and representativeness of data), model structure uncertainty (ability of the model to describe the catchment's response), and parameter uncertainty (adequate values of model parameters). It is difficult to assess the data uncertainty because the magnitude of data errors is often unknown and any attempt to model these deviations is ultimately based on a guess. Model structure uncertainty depends on the choice of the physical or statistical model. It cannot be reduced by increasing information (e.g. the sample size), but only by increasing the knowledge of the process, and by adopting more complex models. The parametric uncertainty is caused by variation in calibration datasets or sampling variability of aparticular model structure. This uncertainty can be included in the predicted variables and presented in a stochastic or probabilistic framework. This provides for the range of possibilities and the likelihood that a given prediction will occur. The quantification of these uncertainties is important for practical decision making. When sufficiently large sets of examples (training patterns) are available, the sampling variability in weights can be approximated by bootstraps (Stone, 1974). Bootstrap technique (Efron and Tibshirani, 1993) has been successfully used in hydrological modelling. The bootstrap is a computational procedure that uses intensive resampling with replacement, in order to reduce uncertainty. In addition, it is the simplest approach since it does not require the complex computations of derivatives and Hessian-matrix inversion involved in linear methods or the Monte Carlo solutions of the integrals involved in the Bayesian approach (Dybowski and Roberts, 2000).

Although several studies indicate that the data-driven models have proven to be potentially useful tools in hydrological modelling, two of the main issues that needs to be further explored before these models gain wider acceptability by researchers and practitioners are: (i) addressing the issues of nonlinearity and non-stationary in NN modeling in a more effective way by considering them on a single platform, and (ii) identifying effective ways for assessing uncertainty in data-driven models, which in turn contributes to improving the reliability of such models.

The present study makes an attempt to address these issues to improve the credibility of the data-driven models among researchers and practitioners. Earlier studies on NN modeling have not given much emphasis on ensemble modeling and uncertainty assessment in hydrologic forecasting for different lead times. The combined strength of bootstrap method with ANN for hydrologic forecasting has not been explored. Also, the uncertainty assessment for longer lead times has not been carried out.

Keeping the above in view, the present research is undertaken with the following objectives:

2. To explore the potential of flood forecasting using ANN in Kulsi River Basin

3. To quantify the uncertainty in flood forecasts using bootstrap technique.

2.0 REVIEW OF LITERATURE

This chapter presents a detailed review of the major issues related to river flow forecasting using neural networks and bootstrap resampling. A brief account of the different studies on river flow forecasting is also presented. Finally, a critique of the review of literature is presented

2.1.Neural Networks

Black box models in the form of neural networks (NNs) have gained momentum in last few decades for river flow forecasting. The FFBP (Feed Forward Back Propagation) is the most popular NN training method advocated in water resources literature. The major advantage of the feed-forward backpropagation NN (FFBP-NN) is that it is less complex than other NN model such as radial basis function (RBF) and has a similar nonlinear input-output mapping capability (Coulibaly and Evora, 2007). Another type of NN model, generalised regression neural network (GRNN) model has also been used successfully for nonlinear input-output mapping but the model has its own advantages and limitations. GRNN model has some advantages of being less sensitive to initial weights and do not produce negative values compared to the FFBP-NN model but some studies show mixed performance of FFBP-NN and GRNN models.

FFBP-NN models (usually called NN models) are known to have several dozens of successful applications in river basin management and related problems. The NN models are widely used in river basin management and are very popular compared to other data driven techniques. Therefore, in this study the NN model is used as it has been increasingly used in rainfall-runoff modelling and river flow forecasting. Substantial literature on NN models have been reported in ASCE (2000a,b). Some of the earlier studies using NN models are reviewed and presented below.

Halff et al. (1993) applied a three layer NN model to portray hydrographs recorded by the United States Geological Survey (USGS) at Bellvue, Washington. They used observedrainfall hyetographs as inputs. Since then, several studies for rainfall-runoff modelling using NN models have been carried out.

Hjelmfelt and Wang (1993) developed a NN model based on the unit hydrograph theory. Using linear superposition, a composite runoff hydrograph for a watershed was developed by the appropriate summation of a unit hydrograph ordinate and the corresponding runoff excesses. NN model was trained to predict runoff from a small watershed in central Missouri. The number of neurons in the input and hidden layers were kept constant. The results obtained showed that a NN model can be constructed successfully to replicate the unit hydrograph.

Zhu et al. (1994) predicted the flood hydrograph in Butter Creek, New York. Predictions with NN model outside the range of training data showed poor results. For longer lead time forecasts, NN model performance deteriorated. The results obtained using NN models were able to perform better than the autoregressive moving average models (ARMA).

Bonafe et al. (1994) analysed the performance of a NN model for forecasting daily mean flow of the upper Tiber river, Italy. The previous day discharge, daily precipitation, daily mean temperature, total rainfall of the previous five days and mean temperature of the previous ten days were selected as NN model input. They concluded that the multi layer NN models are able to yield better performance than ARMA models.

Smith and Eli (1995) applied a back propagation NN model to predict peak discharge and time to peak for a hypothetical watershed of 5×5 grid cells including a drainage pattern to test NN models of rainfall-runoff processes. Linear or nonlinear reservoir models generated dataset for training and testing. Moreover, spatial and temporal distribution of rainfall was incorporated into the NN model. Discharge series were modelled using a discrete Fourier series fit of the rainfall hyetograph.

Tokar and Johnson (1999) reported that a NN model can yield higher training and testing accuracy compared to regression and simple conceptual models. Their aim was to predict daily discharge for the Little Patuxent River, Maryland. Daily precipitation, temperatureand snowmelt equivalent served as inputs in their study. It was found that the selection of training data had a great impact on the accuracy of the prediction. The authors trained and tested their NN model with wet, dry and average-year data, as well as a combination of these in order to illustrate the impact of the training series on network performance. The NN model that was trained on wet and dry year data had the highest prediction accuracy. The length of the training data had a much smaller impact on network performance compared to the type of training data.

Zealand et al. (1999) showed the potential of NN model for short term forecasting of stream flow. They explored the capabilities of NN model to forecast stream flow and compared their performance to conventional approaches. NN models were examined for sensitivity analysis with different type of input dataset as well as number and size of hidden layers. The NN approach was used for a part of the Winnipeg river in northwest Ontario, Canada. Forecasting was conducted on a catchment area of approximately 20,000 km², using quarter monthly time intervals. A very close fit was obtained during the training phase and the developed NN model consistently outperformed a conventional model during the verification (testing) phase for all the four forecast lead-times considered.

Tokar and Markus (2000) compared NN models with traditional conceptual models for predicting watershed runoff as a function of rainfall, snow water equivalent and temperature. NN models were applied to model watershed runoff in three basins with different climatic and physiographic characteristics. In the Fraser river, monthly stream flow was modelled with NN model and compared to a conceptual water balance model. The NN technique was used to model the daily rainfall-runoff process and was compared to the Sacramento soil moisture accounting model in the Raccoon river watershed. The daily rainfall-runoff process was also modelled in the Little Patuxent river basin with NN model. The training and testing results were compared to those of a simple conceptual rainfall-runoff model (SCRR). The results showed that NN models are powerful tools for modelling the precipitation-runoff process.

In an attempt to overcome the limited extrapolation capacity of NN models, Imrie et al. (2000) used different activation functions for the output layer of NN models. They showed that a cubic polynomial function performs better than linear and sigmoid functions for their testing dataset. Further, the results showed that the use of different output layer transfer functions do not have any impact on the quality of training.

Hu et al. (2001) tried to enhance their multi-layer NN based river stage modelling approach by training various NN models for different parts of the observed flow. Therefore, they divided the flow spectrum in low, medium and high flows. This approach has advantage of the closer distance of sampling points each of the NN model is trained on. This consequently leads to better training and generalisation performance of the three networks. They made use of very limited data for training, hindering the application potential for the forecast over the entire range of flows.

Cigizoglu (2003) used ARMA models to generate synthetic time series. These data were incorporated into the training database of NN model to increase the predictive ability. The method was applied to the monthly mean river flow data of a station in Turkey with good results. The extension of input and output datasets in the training stage improved the accuracy of forecasting based on NN model. Introducing periodicity components in the input layer also increased the forecasting accuracy of NN model.

Laio et al. (2003) compared two different nonlinear models, nonlinear prediction (NLP) and NN for multivariate forecasting of the water stages of river Tanaro, Italy. They reported very good results for both the methods. However, nonlinear prediction performed slightly better for short lead times (1-6 h) while the situation was reversed for longer forecast horizons.

Anctil et al. (2004) studied the impact of the length of observed records on the performance of multiple-layer perceptrons (MLPs) and compared their results with those of a parsimonious conceptual model equipped with an updating scheme. Both models were assessed for 1 day ahead stream flow predictions. Ninety-two different model scenarios were obtained for 1, 3, 5, 9 and 15 year time sub-series created from a 24 year training dataset, shifting by a 1 year sliding window. All the model scenarios were verified against the same 7 year dataset. The results revealed that MLP stream flow mapping was efficient as long as wet weather data were available for the training. The physical knowledge in the conceptual models allowed them to make much better use of 1 year training sets than the MLPs. However, longer training sets were more beneficial to the MLPs than to the conceptual model. Both types shared good performance evenly for 3 and 5 year training dataset, but MLPs did better whenever the training dataset was dominated by wet weather. The MLPs continued to improve for input vectors of 9 years and more, which was not the case with the conceptual model.

Castellano-Mendez (2004) compared a forecast algorithm based on the ARMA concept (Box and Jenkins, 1976) with a multi-layer NN based forecast strategy. For daily single step runoff predictions, the nonlinear multilayer NN model performed better than the linear ARMA approach.

Bruen and Yang (2005) used functional networks which were introduced as an alternative for real time flood forecasting. They applied two types of functional network models, separable and associative networks to forecast discharge for different lead times. They compared their results with a conventional NN model, an ARMA model and a simple baseline model in three catchments. Results showed that functional networks are flexible and comparable in performance to NN model. Their results were obtained with only the simplest structure of functional networks and they recommended that the use of complex forms of functional networks might further improve the forecast.

Hettiarachchi et al. (2005) tried to improve the forecast ability of their multi-layer NN model by incorporating an estimated maximum flood (EMF) in the training dataset. This led consequently to better testing results when the EMF hydrograph was contained in the testing dataset. However, this evaluation strategy had little significance as it focused on just one sampling point (the EMF) ignoring the range of all possible flows between the observed and the EMF. The NN model evaluated only five input vectors i.e. two antecedent flows, the actual rainfall and two antecedent rainfall features.

Mutlu et al. (2008) reported on an evaluation of the use of NN models to forecast daily flows at multiple gauging stations in Eucha Watershed, an agricultural watershed located in north-west Arkansas and north-east Oklahoma. Two different NN models, the MLP and the RBFNN models were developed and their abilities to predict stream flow at fourgauging stations were compared. Results revealed that both the models performed satisfactorily for flow predictions at multiple gauging stations, however, the MLP model outperformed the RBFNN model.

Demirel et al. (2009) compared the performance of soil and water assessment tool (SWAT) and NN models for daily flow forecasts of the Pracana basin in Portugal. The NN model was found to be more successful than the SWAT in relation to better forecast of peak flow. The results of this study showed that NN models can be powerful tools in daily river flow forecasts.

Jain and Kumar (2009) presented the results of a study aimed to determine if NN model learn the underlying physical sub-processes during training. This was achieved using simple qualitative and quantitative techniques. The results obtained indicated that the number of hidden neurons determined during training for a particular dataset correspond to certain sub-processes of the overall physical process being modelled. It was found that the time scale of the data employed has an effect on optimum NN model architecture and knowledge extracted.

Akhtar et al. (2009) used NN modelling approach in the Ganges river basin for daily discharge forecasting at Hardinge Bridge, close to its entry point from India into Bangladesh. They presented an analysis of the flow length and travel time as a basis for pre-processing remotely sensed (satellite) rainfall data. This pre-processed rainfall was used together with local stream flow measurements of previous days as input to NN model. The NN model showed its ability to forecast discharge upto 3 day lead time with an acceptable accuracy. Within this forecast horizon, the influence of the pre-processed rainfall was marginal because of dominant influence of strongly auto-correlated discharge inputs. For forecast horizons upto 7 to 10 day, the influence of the pre-processed rainfall was noticeable, although the overall model performance deteriorated.

Kentel (2009) found that even though it is not straightforward to classify an event as regular or extreme with many input parameters, the extreme event can be identified by clustering techniques. He tested the fuzzy c-means (FCM) clustering algorithm and found that the FCM clustering satisfactorily classified input vectors into ‘‘regular event” or‘‘extreme event” classes, indicating high potential of the clustering techniques for water resources planning and management problems.

Wu et al. (2010) proposed modular artificial neural network (MANN) model for daily and monthly rainfall forecasting. MANN model was developed by partitioning the training data into three clusters by the FCM technique, and then each subset was approximated by a single NN model. The final output of the model resulted directly from the output of one of three local models. MANN was compared with three benchmark models, viz. NN, K-nearest-neighbours (K-NN) and linear regression (LR). Results of models without preprocessed inputs indicated that MANN model performed the best among all four models. When all the four models were coupled with three data-preprocessing techniques viz. moving average (MA), principal component analysis (PCA) and singular spectrum analysis (SSA), results showed that improvement of model performance generated by SSA was considerably good, whereas those of MA or PCA were marginal. MANN model coupled with SSA was found to be the best compared to remaining models, especially for daily rainfall forecasting.

Mount and Abrahart (2011) suggested the use of linear models as standards against which artificial Intelligence (AI) techniques are to be tested, indicating the degree to which input-output relationship being modelled require the application of complex AI modelling techniques.

2.2. Bootstrap Method

The bootstrap (Efron, 1979; Efron and Tibishirani, 1993) is a computational procedure that uses intensive resampling with replacement to reduce uncertainties. The aim of re-sampling is to mimic the random component of a process and to reduce variance through averaging over numerous different partitions of the data. It is commonly used to estimate confidence intervals, but it can also be used to estimate bias and variance of an estimator or calibrate hypothesis tests.

Bootstrapping generates multiple versions of predictor to obtain an aggregated predictor. The aggregation averages over the versions when predicting a numerical outcome by making bootstrap replicates of the learning sets and using these as new learning sets (Breiman, 1996). Bootstrapping has been implemented in diverse environment related applications such as toxicology (Bailer and Oris, 1994), fisheries surveys (Smith, 1997), groundwater and air pollution modeling (Archer and Giovannoni, 1998), hydrology (Fortin et al., 199), spatial point patterns (Solow, 1989)and ecological indices (Dixon, 2001).

The use of non-parametric bootstrap approaches in hydrological modeling is on the increase. Documented applications range from estimating means, confidence intervals, parameter uncertainties and network design techniques (e.g. Cover and Unny, 1986; Tasker, 1999; Woo, 1989; Moss and Tasker, 1991; Zucchini and Adamson, 1989; Di Stefano et al., 2000) to the adoption of more complicated block based methodologies that endeavour to maintain temporal dependence or spatial covariance (e.g. Lall and Sharma, 1996; Vogal and Shallcross, 1996; Sharma et al., 1997; Tasker and Dunne, 1997; Srinivas and Srinivasan, 2000, 2001). The application of bootstrap methodologies to build neural networks is also the subject of current research.

2.2.1Bootstrap-based ANN Models (BANNs)

There are two natural paths for randomness to enter a neural model-building operation: through different choices about splitting the data, or through different choices about network initialization, architecture and training. Either path or both paths together can be bootstrapped. The neural bootstrap has been used to perform bootstrap aggregation also known as bagging of multi-model ensembles to produce averaged outputs and a more stable solution (Hsieh and Tang, 1998; Tang et al., 1998) and bootstrap assessment of multi-model multi-data solutions to establish the influence of different components (LeBaron &Weigend, 1998). More sophisticated neural bootstraps have also been used to estimate confidence bounds for network outputs (Efron and Tibshirani 1993) and for bootstrapping residuals to evaluate forecasting power (Weigend et al., 1992) or to obtain error bars on iterated time series predictions (Connor, 1993).

Comparison of a continuous single-model bootstrap with conventional stopping conditions applied to neural networks was carried out by Abrahart (2003) for rainfall-runoff forecasting of Upper River Wye basin in Central Wales. The results obtained in the study demonstrated the marginal improvements in terms of greater accuracies and better global generalizations and substantial advantages in the form of automation and diagnostic capabilities from the bootstrap based neural networks. Ensemble neural network (ENN) approach using bagging was also compared with single neural network (SNN) for rainfall-runoff modeling in a study undertaken by Jeong and Kim (2005). Both ANN models used the early stopping method to optimize generalization performance during training. The ANN models were applied to make

1-month ahead probabilistic forecasts for inflows to the Daecheong multipurpose dam in Korea. The results of the study revealed that ENN was less sensitive to the input variable selection and the number of hidden nodes than SNN. ENN in general, produced smaller RMSEs than the corresponding SNN, signifying that the ENN reduces the generalization error more efficiently compared to the SNN.

The performance of a bootstrapped artificial neural network (BANN) wascompared with maintenance of variance extension (MOVE) and a modified drainage area ratio (DAR) for synthetic flow generation with a small data sample in a study undertaken by Jia and Culver (2006). The bootstrap method was used to estimate the generalization errors of neural networks with different structures and to construct the confidence intervals for each flow prediction. Comparisons were investigated with respect to a case study at Buck Mountain Run, Albemarle, VA, USA, with various small data sets of flow observations and flow predictors spanning different time periods. The results of the study showed that BANN outperformed MOVE and DAR for the predictions of the low and medium flow.

Han et al. (2007) studied the uncertainties involved in real time forecasting using a NN model. They proposed a method to understand the uncertainty in NN hydrologic models with the heuristic that the distance between the input vector at prediction and all the training data provide a valuable indication on how well the prediction would be. They concluded that for long term predictions, the NN model showed superior performance but that was only probabilistic depending on how the calibration and test events were arranged. He showed that it is essential to improve the understanding about the uncertain nature and hydrological characteristics of the NN model in flood forecasting for its use as effective practical tool. However, their method did not quantify the uncertainty of the model parameters or the predictions.

Srivastav et al. (2007) proposed a method of uncertainty analysis for NN hydrological models which was based on bootstrap technique. They developed a NN model for forecasting the river flow at 1 h lead time. The results revealed that the proposed method of uncertainty analysis is very efficient and can be applied to a NN based hydrologic model.

Sharma and Tiwari (2009) developed bootstrap based artificial neural network (BANN) model for hierarchical prediction of monthly runoff and found that the performance of BANN model was much better compared to SNN model.

3.0 THEORETICAL CONSIDERATIONS

This chapter deals with the theoretical background of neural networks (NNs) and bootstrap resampling that are used for the development of different NN models.

3.1 Neural Networks

Neural networks (NNs) are information processing systems composed of simple processing elements (neural units) linked by weighted synaptic connections (Rumelhart and McClelland, 1986; Muller and Reinhardt, 1991). They reconstruct the complex nonlinear input/output relations by combining multiple simple functions, by analogy with the functioning of the human brain. From the mathematical point of view, NN can be considered to be a multivariate nonlinear nonparametric statistical method (White, 1989; Ripley, 1993). This approach is fast and robust in noisy environments, flexible in the range of problems it can solve, and highly adaptive to newer environments. Owing to these established advantages, NNs currently have numerous real-world applications, such as time series prediction, rule-based control and rainfall-runoff modelling (Jain et al., 1999).

The multilayer feedforward NN consists of a set of sensory units that constitute the input layer, one or more hidden layers of computation nodes and an output layer of computation nodes. The input signal propagates through the network in a forward direction, layer by layer. These NNs are commonly referred to as multilayer perceptrons. A typical three-layer feedforward NN is shown in Fig 3.1. The notation I_i is the input value to node i of the input layer, V_j is the hidden value to node j of the hidden layer, and O_k is the output at node k of the output layer. An input layer bias term I_0 =1 with bias weights w_{j0} and an output layer bias term V_0 =1 with bias weights w_{k0} are included to permit adjustments of the mean level at each stage. The mathematical form of a three-layer feedforward NN is given as:

$$O_k = g_2\left[\sum_j w_{kj}\ g_1\left(\sum_i w_{ji}\ I_i + w_{jo}\right) + w_{ko}\right] \tag{3.1}$$

Where two sets of adjustable weights, w_{ji} controls the strength of the connection between the input node i and the hidden node j, and w_{kj} controls the strength of the connection between the hidden node j and the output node k. g_1 and g_2 are activation functions for the hidden layer and the output layer, respectively. The activation function is usually selected to be a continuous and bounded nonlinear transfer function, for which the sigmoid and hyperbolic tangent functions are commonly used (e.g. Hsu et al., 1995; Haykin, 1999; Govindaraju and Rao, 2000).

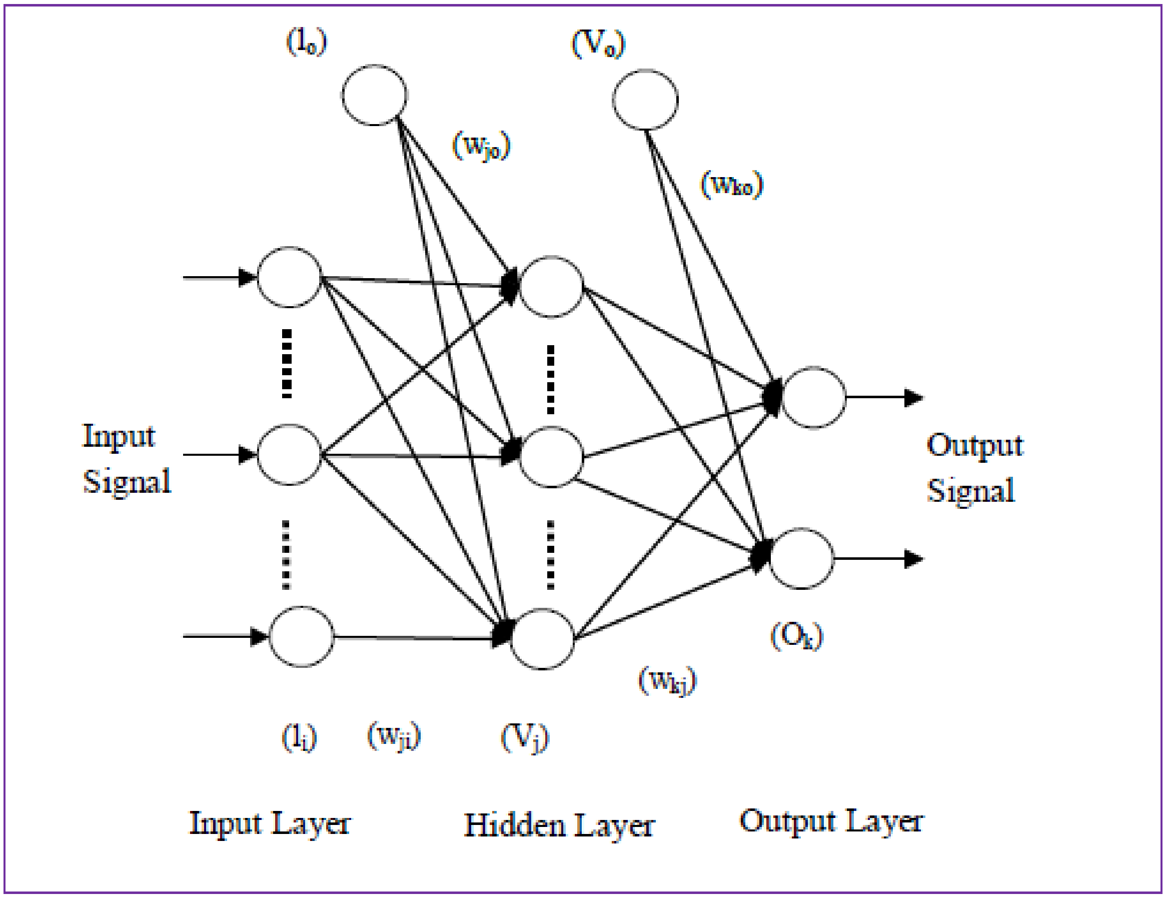


Fig 3.1 A Basic Overview of NN Topology

3.1.1 Neural units

A simple neural model is presented in Fig. 3.2. From mathematical point of view information processing within a neuron involves two distinct mathematical operations: (a) Synaptic operation that assigns a relative weight (significance) to each incoming signal according to the past experience (knowledge) and (b) Somatic operation that provides various mathematical operations such as aggregation, thresholding, nonlinear activation, and dynamic processing to the synaptic inputs. On the basis of order of synaptic operations used, different neural units can be formulated (Gupta et al., 2003; Gupta et al., 2009).

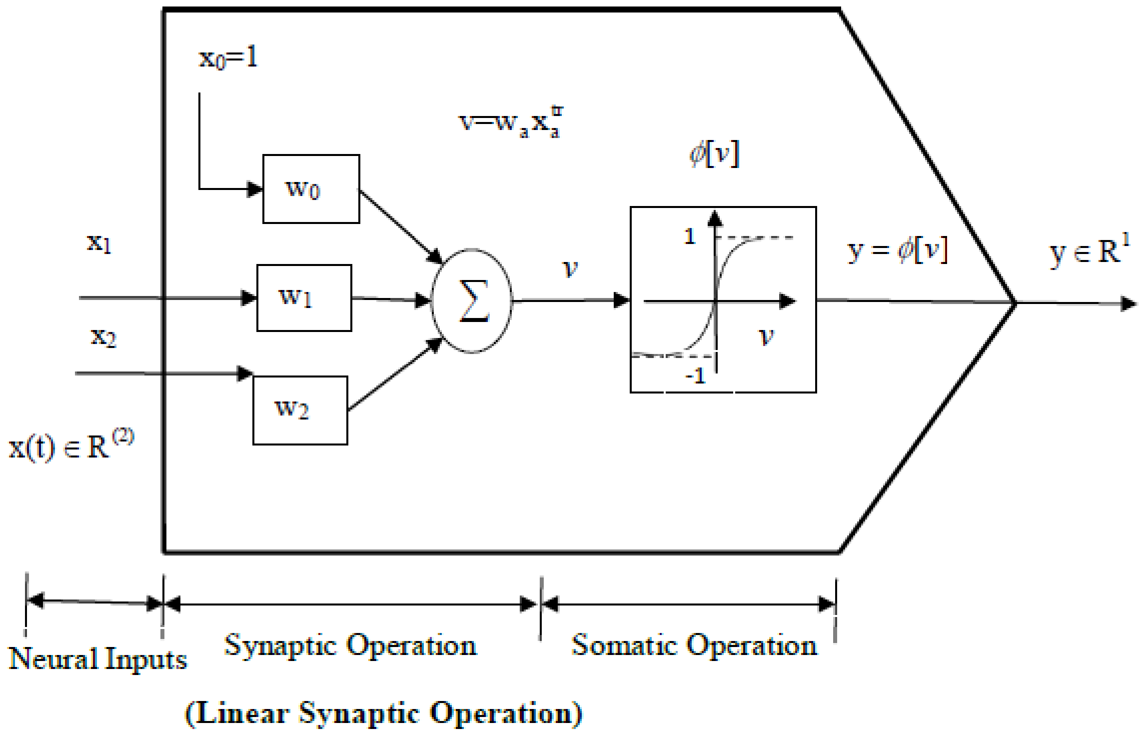


Fig 3.2A Simple Mathematical Neural Unit

3.2 Multilayer Perceptron

In this study, neural network analysis was performed using *Neuropath* software (Minasny and McBratney, 2002). A neural network is an attempt to build a mathematical model that supposedly works in an analogous way to human brain. A network consists of many elements or 'neurons' that are connected by communication channels or 'connectors'. These connectors carry numeric data arranged by a variety of means and organized into layers. The neural networks can perform a particular function when certain values are assigned to the connections or 'weights' between elements. To describe a system, there is no assumed structure of the model, instead the networks are adjusted or 'trained' so that a particular input leads to a specific target output (Gershenfeld, 1999). The mathematical model of a neural network comprises of a set of simple functions linked together by weights. Figure 3.3 shows the schematic representation of the architecture of ANN's. The type of ANN described in the figure 3.3 is called the multilayer perceptron (Nørgaard, 2000).

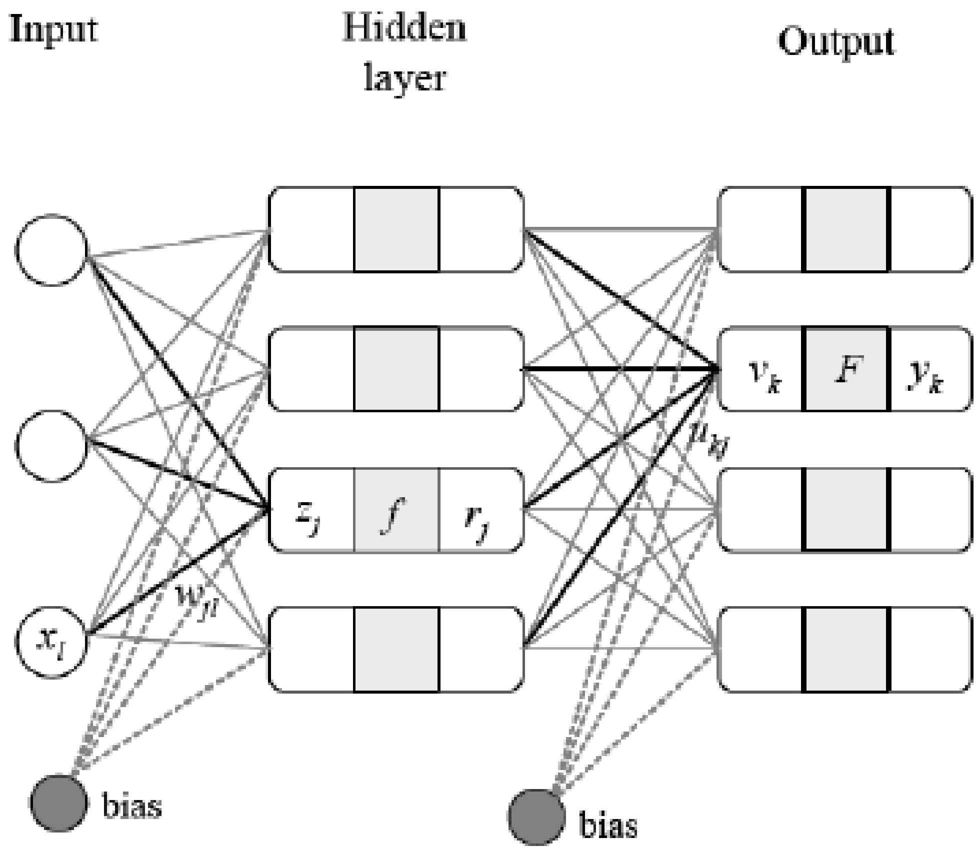


Fig 3.3 Schematic Diagram of Multilayer Preceptron (MLP)

The network consists of a set of input units *x*, output units *y*, and hidden units *z*, which link the inputs to outputs. The hidden units extract useful information from inputs and use them to predict the outputs. Input vector of elements x_l (l = 1…., N_l) is transmitted through network connections that is multiplied by weights W_{jl} of each connection to give the internal activity of each hidden neuron as z_j (j = 1,…,N_h) :

$$z_j = \sum_{l=1}^{N_l} w_{jl} x_l + w_0 \tag{3.2}$$

Where, N_h is the number of hidden neurons and N_l is the number of input source nodes. The hidden units consist of the weighted input (w_{jl}) and the bias (w_0). A bias (w_0) of input equal to 1 that serves as a constant added to the weight. These inputs are passed through a layer of activation function f which produces:

$$r_j = f\left(\sum_{l=1}^{N_l} w_{jl}\, x_l + w_{j0}\right) \tag{3.3}$$

The activation functions are designed to accommodate the nonlinearity in the input-output relationships. The function used in *Neuropath* software is the hyperbolic tangent sigmoid:

$$f(z) = \tanh(z) = 1 - \frac{2}{1+\exp(2z)} \tag{3.4}$$

The outputs from hidden units pass another layer of filters:

$$v_k = \sum_{j=1}^{N_h} u_{kj}\, r_j + u_{ko} = \sum_{j=1}^{N_h} u_{kj}\, f\left(\sum_{l=1}^{N_l} w_{ji}\, x_l + w_{jo}\right) + u_{ko} \tag{3.5}$$

and fed into another activation function (F) to produce output y_k (k = 1,….N_0)

$$y_k = F(v_k) = F\left(\sum_{j=1}^{N_h} u_{kj}\, f\left(\sum_{l=1}^{N_l} w_{jl}\, x_l + w_{j0}\right) + u_{k0}\right) \tag{3.6}$$

The weights are adjustable parameters of the network and are determined from a set of data though the process of training. An adaptive nonlinear least squares algorithm developed by Dennis et al. (1981) is implemented in the *Neuropath* software, which was used for training the networks. The objective of training was to minimize the sum of squares of the residuals between the measured and predicted outputs.

$$O(W,U) = \sum_{i=1}^{N_i} \sum_{k=1}^{N_0} (\widehat{P_{ik}}\ (X_i) - P_{ik})l^2 \tag{3.7}$$

where N_s is the number of datasets, N_0 is the number of outputs, W and U are weights of the hidden and output layer, respectively, $\overline{P_{ik}}$ is the measured output and isthe predicted outputfrom the input vector X

3.3 Bootstrapping

Neural networks have been combined with the bootstrap method in the *Neuropath* software. Efron and Tibshirani (1993) developed Bootstrapping method which manipulates the training datasets in order to generate different models and uses them to obtain an aggregated predictor. The bootstrap procedure involves resampling with replacement, to reduce uncertainties. The aim of resampling is to mimic the random component of a process and to reduce variance through averaging over numerous different partitions of the data.

Neural networks simulations generally converge at the local minima. This results in slightly different predictions every time neural networks are trained, due to the random initializations of the weight matrix. The neural bootstraps have been used to perform bootstrap aggregation (bagging) of multi-model ensembles which produced averaged outputs and a more stable solution. Bootstrap assumes that the training dataset is a representation of the population, and multiple realizations of the population can be simulated from a single dataset. This is done by repeated 'sampling with replacement' of the original dataset of size of N, to obtain B bootstraps datasets, each with size of N. Each bootstrap dataset contains a different data, resulting in B neural network models, all of which may differ slightly. Efron and Tibshirani (1993) suggested B to be between 50 and 200.

Since resampling is done with replacement, each sample has a chance of $1-[(N-1)/N]^N$ (approximately 63 %) to be selected once or multiple times from a bootstrap dataset. Each bootstrap dataset thus contains one or more copies of 63 % of the samples. By calculating averages and standard deviations of B testing results, one obtains robust values of the predicted variable and associated uncertainty estimates for independent data. A model f(x) is fitted to each of the generated bootstrap datasets and bootstrapping estimate $f_{bootstrap}(x)$ is calculated as the mean of each model:

$$f_{bootstrap}\ (x) = \frac{1}{B}\sum_{b=1}^{B} f(x) \tag{3.8}$$

Bootstrapping also known as bagging approach of generating different models and aggregating them to produce an estimate, which has been found to increase the accuracy of neural networks (Breiman, 1996).

4.0 STUDY AREA

4.1 Location

The Kulsi basin,a part of the Brahmaputra sub-basin is situated on the south bank of the mighty river Brahmaputra (Fig 4.1). This sub-basin spreads in the Kamrup District of Assam as well as west Khasi hills and Ribhoi district of Meghalaya. It is located between latitude 25°30'N to 26°10'N and longitude 89°50'E to 91°50'E with an altitude between 100 m to 1900 m above msl.

The river Kulsi drains out a total area of 2806 sq. km. within the Kamrup District of Assam as well as west Khasi hills and Ribhoi district of Meghalaya.

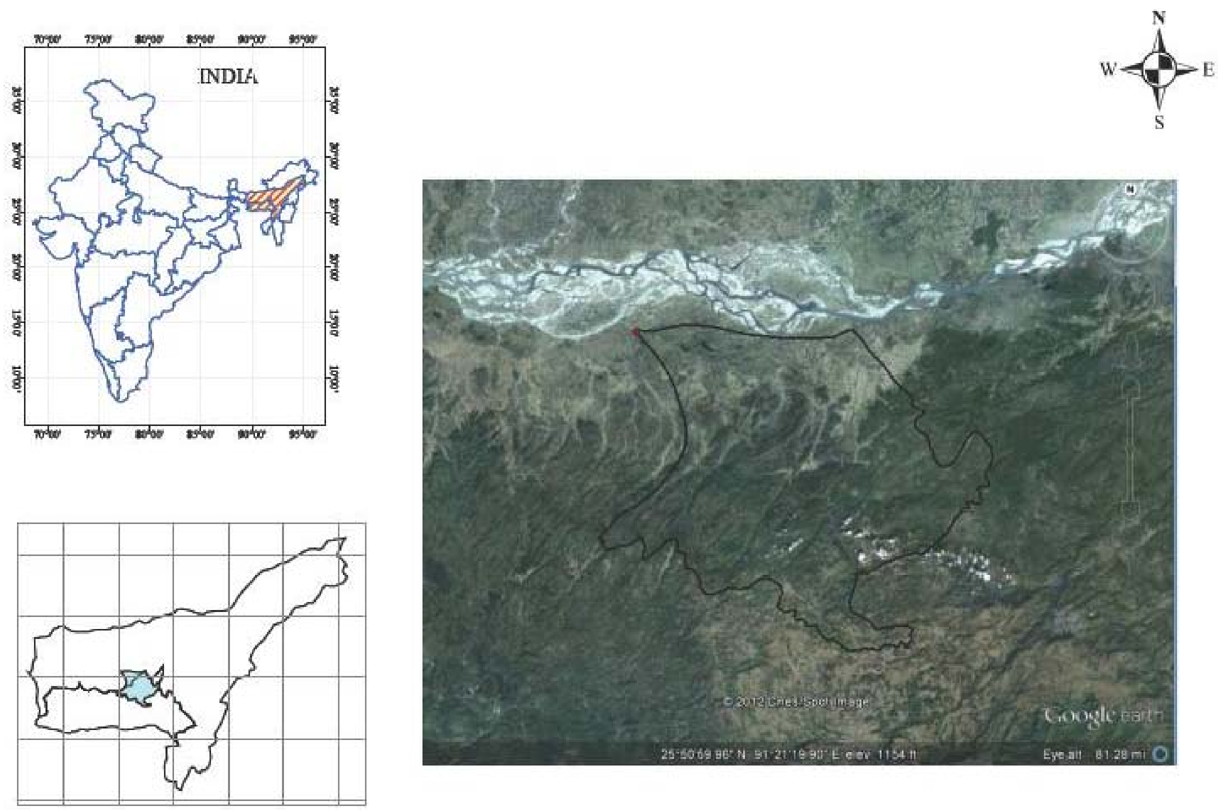


Fig.4.1 Location of Kulsi River Basin

4.2. River System

The Kulsi River is formed by the joining of three rivers namely Khari, Um Krishniya and Um Siri, all of which originate from the west Khasi Hills ranges and flows down north and joins the Brahmaputra (Fig 4.2). This hill range is covered with evergreen forests and gets high rainfall during the summer months. The river Um Krishniya is the centrally flowing river- with Um Siri joining it on the left bank and Um Khri joining it on the right bank. Um Krishniya and Um Sirioriginate almost from the same place, which is having the altitude around 1550m and Um Khiri originate from on area further east, which have got elevation around 1600m. It is interesting to note that all the three rivers originate from more or less the same latitude of 25035'. All the three rivers are joined by innumerable no. of small hilly streams and rivulets till they join together and flow down north by the name Kulsi.

The river Khiri and Um Krishniya join together after they flow respectively for a distance of about 85 km. and 47 km. After joining, the river flows by the name Khri for a distance about 15 km. when it is joined by UmSiri, which flows for a distance of about 52 km., till it joins here. After this point the river flows almost straight south for a distance of about 20 km. by the name Kulsi near to village Kulsi wherein it bifurcated into two branches of by the western side of Kulsi reserve forest and the other the eastern side of it and both flow by the name Kulsi one is eastern Kulsi and the other central Kulsi. The central Kulsi again bifurcates into near village Hatigarh, and the left flowing one flows by the name Kharkhari and the other flows by the original name Kulsi. After these branches off, the river Kulsi enters into the alluvial plain (Flood plain of the river Kulsi and Brahmaputra) and gets shallow and meanders. The eastern most channel (Kulsi) again branches of into 3 channels and again join together before crossing the N.H.- 37, near Kukurmara. In this reach the river is joined by another two small river Batha and Bahwa. After crossing N.H way it takes a complete western turn and flow parallel to the Brahmaputra and joins the other branch of the Kulsi (western Kulsi), which crosses the N.H. at Chaygaon, near village Satghariapara and both flows parallel to river Brahmaputra and meets the other branch of the Kulsi i.e. the Kharkhari near village Chamaria and flows west parallel to the Brahmaputra by the name Jaljali till it joins the Brahmaputra near Bahati. After the bifurcation of the river Kulsi, it is seen that the river takes westerly swing and flows paralleled to Brahmaputra, keeping a distance of 6 to 12 km. from the Brahmaputra and in this reach, it intercepts all the south flowing rivers coming down from the western Khasi Hills and Garo Hills ranges.

The drainage pattern of the Kulsi River systems is often found to be trellis and rectangular type in the upper catchment region, which developed due to structural control on the drainage network.

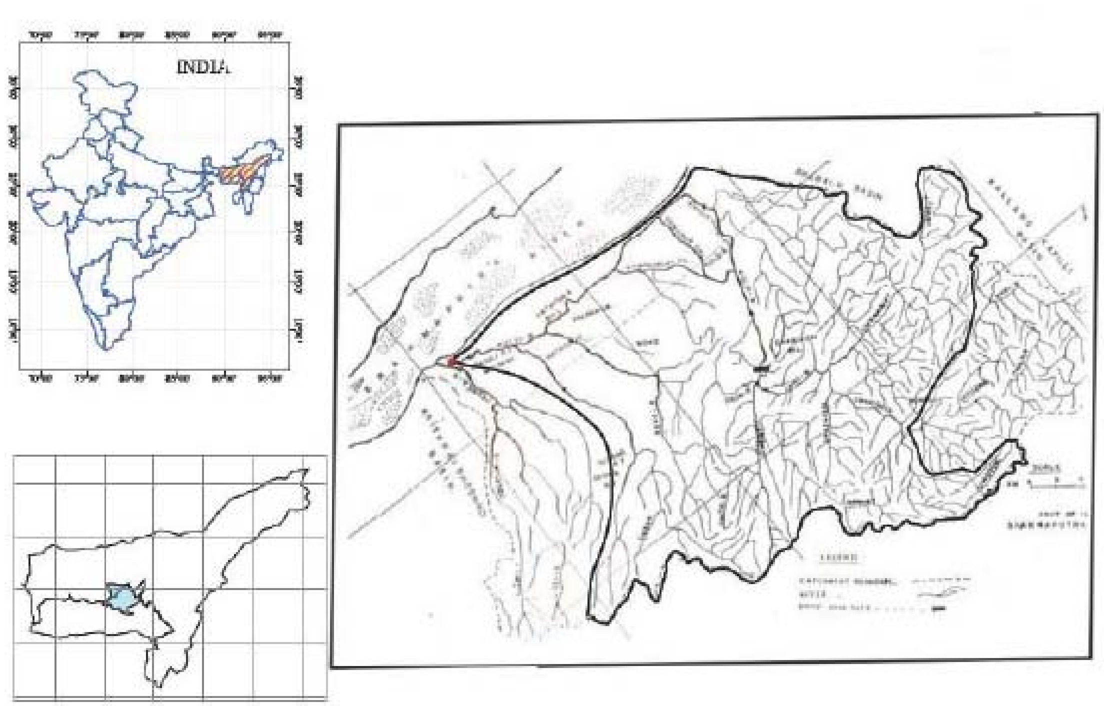


Fig.4.2 Map showing main river system of Kulsi Basin

4.2.1 The tributaries of Kulsi

The Kulsi River receives large number of tributaries in its various reaches.In the upper reach, it is joined by Krishniya and Umsiri on the left bank. In the middle reach its main tributary is Batha while in the flood plain reach it is joined by Boko and Singra. The catchment areas of the main tributaries of Kulsi are as follows (Table.4.1):

Table 4.1 Tributaries of Kulsi and their Catchment Area

Sl. No	Name of the tributaries	Catchment Area (Sq. Km)
1	Krishniya	70 sq.km
2	Umsiri	295 sq.km
3	Batha	285 sq.km
4	Boko	250 sq.km
5	Singra	380 sq.km

4.3. Topography and Basin Characteristics

The topography of Kulsi River Basin is shown by Digital Elevation Models (DEM) downloaded from Shuttle Radar Topography Mission (SRTM) of 90m resolution which is available freely in the internet. Figure 4.3 shows the DEM of the Kulsi River Basin.

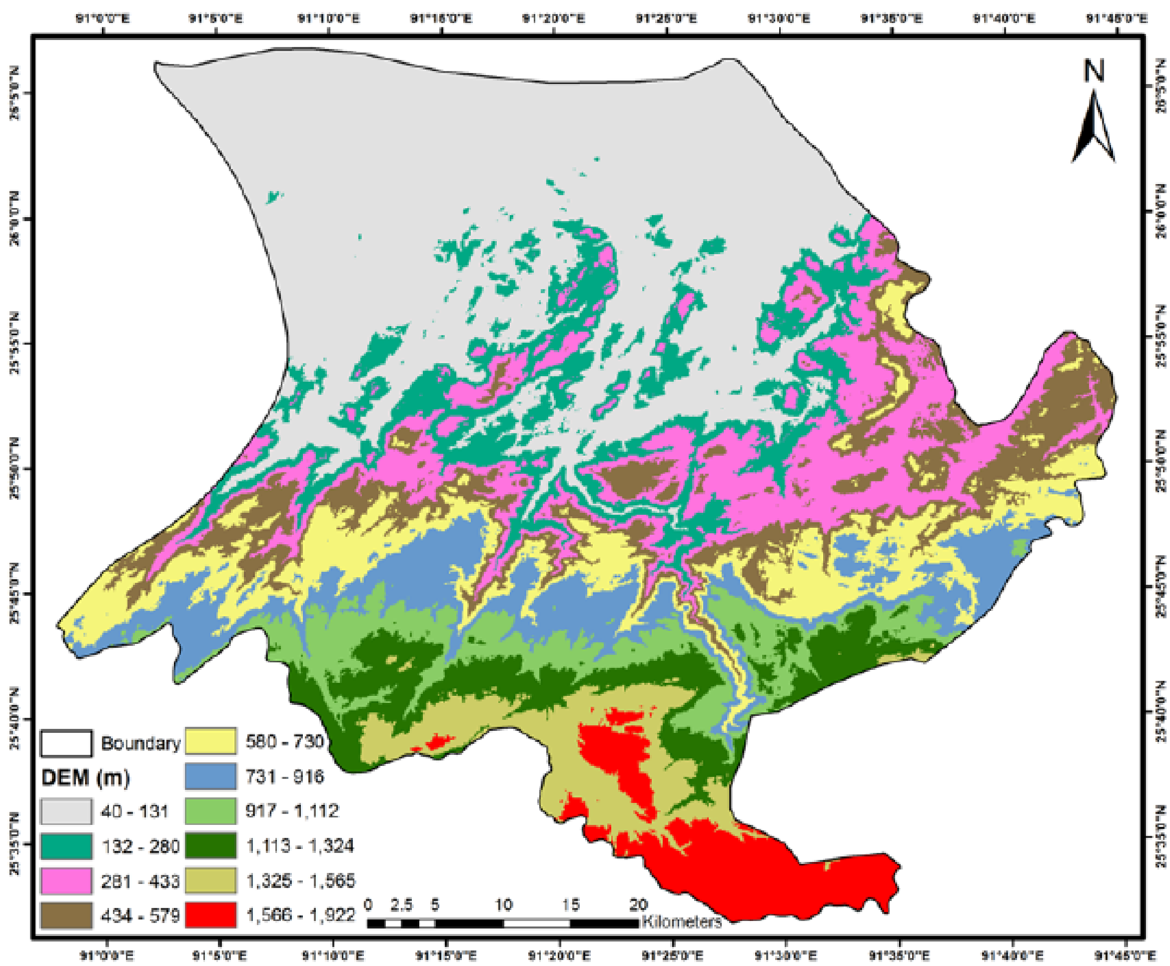


Fig.4.3 Digital Elevation Model of Kulsi Basin

Different parameters like Slope (in percent), Aspect and flow direction can be derived from the DEM using GIS Platform**.**

The area that the Kulsi River drains can broadly be divided into three reaches

(i) The Upper Khasi hill reach

(ii) The middle reserve forest reach

(iii) The alluvial or flood plain reach

The Upper Khasi hill ranges of the catchment extend from the origin of the river Kulsi to Ukium (Assam Meghalaya border) and this reach lies entirely in the west khasi hill ranges, with the general altitude varying from 150 m to 1900 m. the whole area consists of series of hill range with intermittent plain areas. The upper reaches of the region is covered with evergreen and deciduous forest. However, most of the area represents degraded forest and in many places are covered by scrubs . No area in the reach remains under snow cover.

The middle reserve forest reach region consists of areas along the southern bank of river Brahmaputra. The region near Kulsi village, the river Kulsi branches off into three channels and all of them join together after flowing a distance of few kilometers downstream. The middle reserve forests reach consists of two reserve forests namely Borduar and Pantan reserve forest running parallel along the river from Ukium to Kulsi village with the Borduar reserve forest on eastern bank and Pantan reserve forest on the western bank. The river in this reach has got a very narrow valley running between these two reserve forests. The eastern part of the Borduar reserve forest consists of comparatively plain areas with the famous Chandubibeel located therein. No tributary joins the river in this reach.

The alluvial or the plain reach consists of the plain areas along the southern bank of river Brahmaputra. Almost half of this reach is affected by the flood of the River Kulsi and theBrahmaputra. Just at the starting of this reach i.e. near village Kulsi, the river Kulsi branches off into three channels and all of them join together after flowing a few kms and outfalls into the Brahmaputra near Barak. This area is fully cultivable area and sufficiently densely populated

The bed slops of river Kulsi at different reaches are as below:

From source to confluence with Um Trisung	1 in 20
From confluence with Um Trisung to confluence with Khari	1 in 44
From confluence with Khari to Ukium	1 in 80
From Ukium to Dam site	1 in 1570

4.4 Temperature

The climate of Kulsi basin, excluding the upper most reach is similar to that of the other districts in Central Assam. The winter is cold and foggy, while the summer is hot and humid. There is no meteorological centre within the catchment for observation of temperature. However the nearest observatories for the basin are at Guwahati, Umiam and Shillong. Based on long term data from these stations it has been observed that the average maximum temperature in this basinvaries between 15^{0}C to 33^{0}C and average minimum temperature varies from 3^{0}C to 12^{0}C.

4.5 Humidity

The climate of this sub-basin is generally very humid. There is no meteorological Centre in the sub-basin for observation of humidity. In Brahmaputra valley, Humidity is high at a place where forest cover and vegetation cover is relatively more. Humidity data is available only at Guwahati nearest to the sub-basin. Guwahati city is in the East and situated on the south bank of Brahmaputra.

Observing the data it can be concluded that in the Kulsi-Deosila basin in general, the humidity is maximum during June, July, August and September when average relative humidity various from 82% to 86% at 1730 hours. March is the driest month with humidity fluctuating from 58% to 64% at 08-30 hrs and 46% to 74% at 17-30 hrs.

4.6 Geology

The geology of the river basin consists mostly of gneiss and sandstones overlain by a deep to moderately deep soil layer. Much of the terrain is rough, rolling to steeply sloping. Under saturated conditions, such a formation is highly conducive to rapid subsurface storm flow. The rock types in the Kulsi basin vary from Precambrian stage to recent. The surface Geological formation is newer alluvium sand, gravel, clay and silt. In Assam part of the basin falls in Kamrup District where two distinct groups of rock units i.e. consolidated and unconsolidated formation of rocks are found.

The unconsolidated formations represented by the alluvial deposits of recent age such as sand, gravel, pebble, silt and clay. In the Meghalaya part of the basin, there are two, three types of formations like the Archaen complex, lower, cordovan rocks, and cretaceous tertiary sediments. The oldest formation of upper tertiary sediment occurs in Garo Hills.

4.7 Land Use Pattern

The basin falls under warm humid sub-tropical climate, which is suitable for growth of semi-evergreen and deciduous forests dominated by Sal Plants. Most of the areas were afforested by Forest Department, Government of Assam with Sal & Teak Plants, Banana Plants, Shrubs etc. Bamboo is one type of dominant economic plant available in all parts of the area. The land use consists primarily of evergreen forest, semi-evergreen forest, moist-deciduous forest, bamboo-thickets, jhum and rolling grasslands. The pattern of land use of the Kulsi basin area as follows:

1. Degraded forest	62.33%
2. Dense forest	15.56%
3. Scrub	15.50%
4. Jhum cultivation	0.50%
5. Agricultural plantation	2.41%
6. Agricultural land	2.37%
7. Rivers	1.04%
8. Water bodies	0.29%

Total 100.00%

False Color Composite (FCC) draped over DEM of Landsat ETM of Kulsi Basin dated 26-10-2006 shows the present Landuse Pattern of Kulsi River Basin (Fig 4.4).

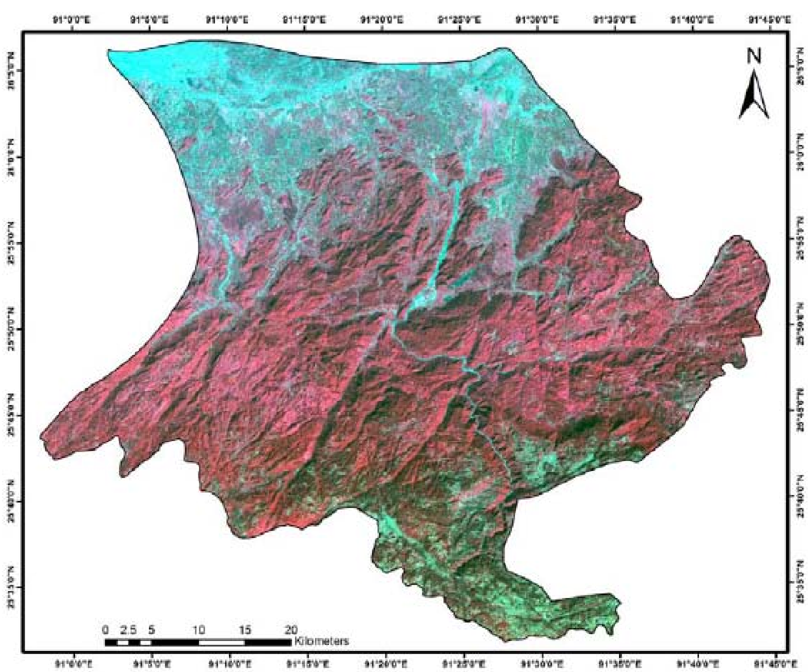


Fig.4.4: FCC draped over DEM of Landsat ETM of Kulsi Basin dated 26-10-2006

4.8 Season

There are four numbers of different seasons occurring in Kulsi-Deosila sub-basin. These are: (1) Winter (2) Summer (3) Monsoon and (4) Autumn or post-Monsoon.

i) **Winter:** The winter season sets in December and ends in February. This is the coldest season. The weather changes due to passage of western disturbance over the region. Low winter rainfall also occurs occasionally.

ii) **Summer:** The summer season begins in March and continues up to May. In this season, occasionally marked in instability develops in the atmosphere and severe thunder storms occur, sometimes preceded by dust raining squalls. Rainfall increases both in quantity and frequency, as the season advances and generally associated with thunder storms and squalls. Hail storms occur sometime in the season specially in hills.

iii) **Monsoon:** The monsoon sets in the last week of May or in early June. It generally occurs due to depression in the Bay of Bengal. Subsequently a series of such depression forming at the head of Bay of Bengal and moving inland; give spells of continuous and moderate to heavy rain over the sub-basin. The monsoon withdraws in the last week of September or first week of October.

iv) **Autumn or Post Monsoon (October to December):** This season begins in October and ends in December**.** There is almost no rain during this period and the climate is neither very cold nor hot.

4.9 Rainfall

Rainfall in the sub-basin occurs mainly due to the monsoon which sets in this region in the first week of June and continues till the 1week of September. About 65% of the total annual precipitation occurs during this period in this sub-basin. After September, the intensity of rainfall gradually decreases. Before the onset of monsoon, sometimes considerable thunderstorm occurs in the region in the month of April and May, resulting in significant rainfall of about 25%. Generally December, January and February are the dry months in the sub-basin when there are only occasional rainfalls.

There are 5 (five) nos. of Rain gauge stations in and around the sub-basin. They are Boko, Chamaria, Mirza and Guwahati Airport. The other one is Ukium, which is maintained by Brahmaputra Board.

5.0 METHODOLOGY

This chapter deals with the methodology used for the development of Bootstrap based Artificial Neural Networks (BANNs). The performance measures used to assess the performance of developed models are also presented. The accuracy of any prediction tool depends largely on the accuracy of datasets with relevant information and the methodology adopted. In modelling the hydrological processes like floods, the causative factors such as climate, soil type, land use, topography, geology and morphological characteristics should be known along with the hydrologic time series data. But, adequate data for a complete and comprehensive analysis are seldom available, leaving no other choice than to rely on the available information. Figure 5.1 shows the flowchart of the methodology adopted in the study.

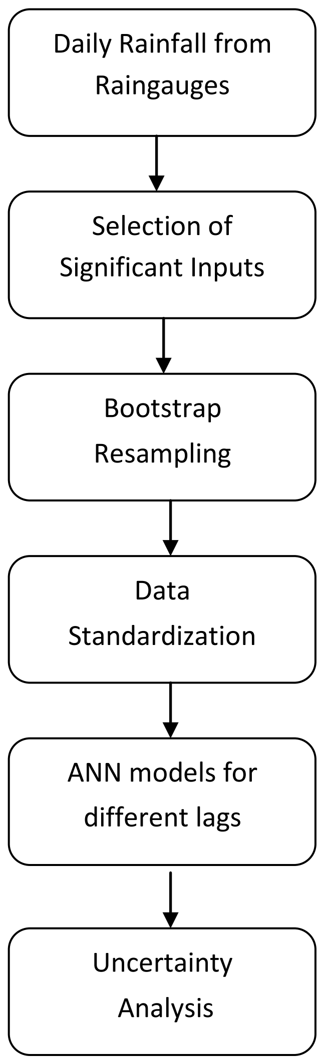


Fig 5.1 Flowchart of methodology adopted in the study

5.1 Data Standardization

The values of daily rainfall and discharge were standardized using the following equation:

$$\frac{x_i - x_{min}}{x_{max} - x_{min}} = x_{new} \tag{5.1}$$

Where x_{i} is the original value of the input, $x_{\mathrm{new}}$is the standardized value of the input, $x_{\mathrm{max}}$and x_{min} are the maximum and minimum values of the input respectively. The values of input vary from 0 to 1 after standardization.

5.2*Neuropath* Software

The software was developed by Budiman Minasny and Alex. B. McBratney of Australian Centre for Precision Agriculture. It was originally developed to fit pedotransfer functions of soil but can be employed for general predictions as well. The interface of *Neuropath* software is shown in figure 5.2

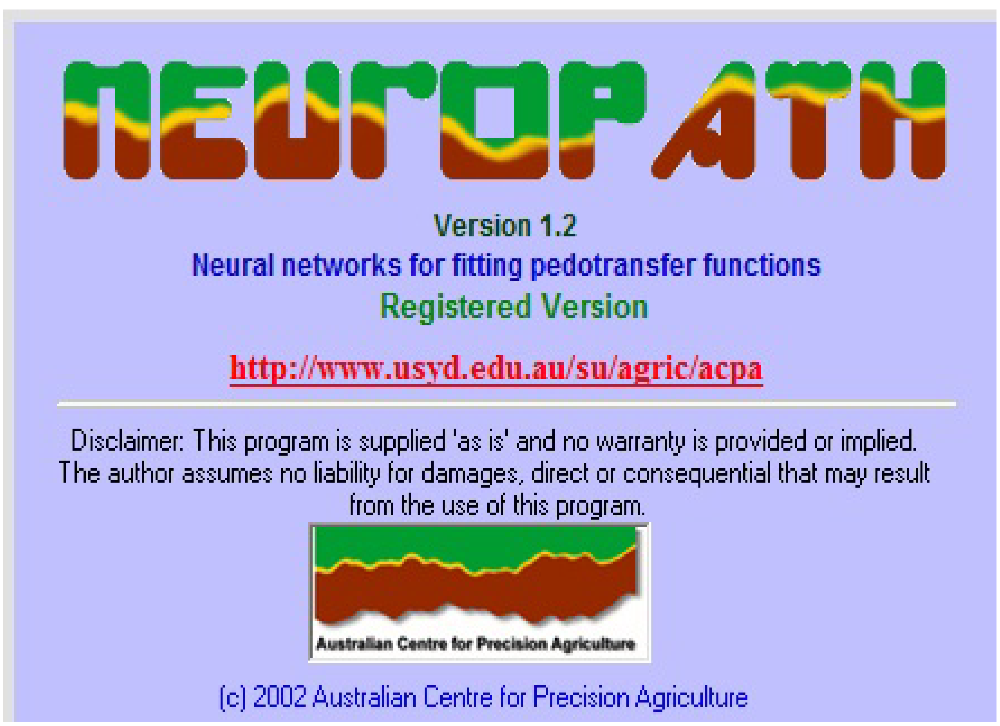


Figure 5.2 *Neuropath* Software interface

The required input in the software is given in ASCII text format. Data is arranged in rows and columns, separated by comma, or tabs or spaces. First row is the header for each column. Heading for each column should be given with no space. First column is the id for the values, followed by the inputs from second column, then the outputs.

5.2.1 Running the Program

The program will start with the screen as shown in figure 5.3

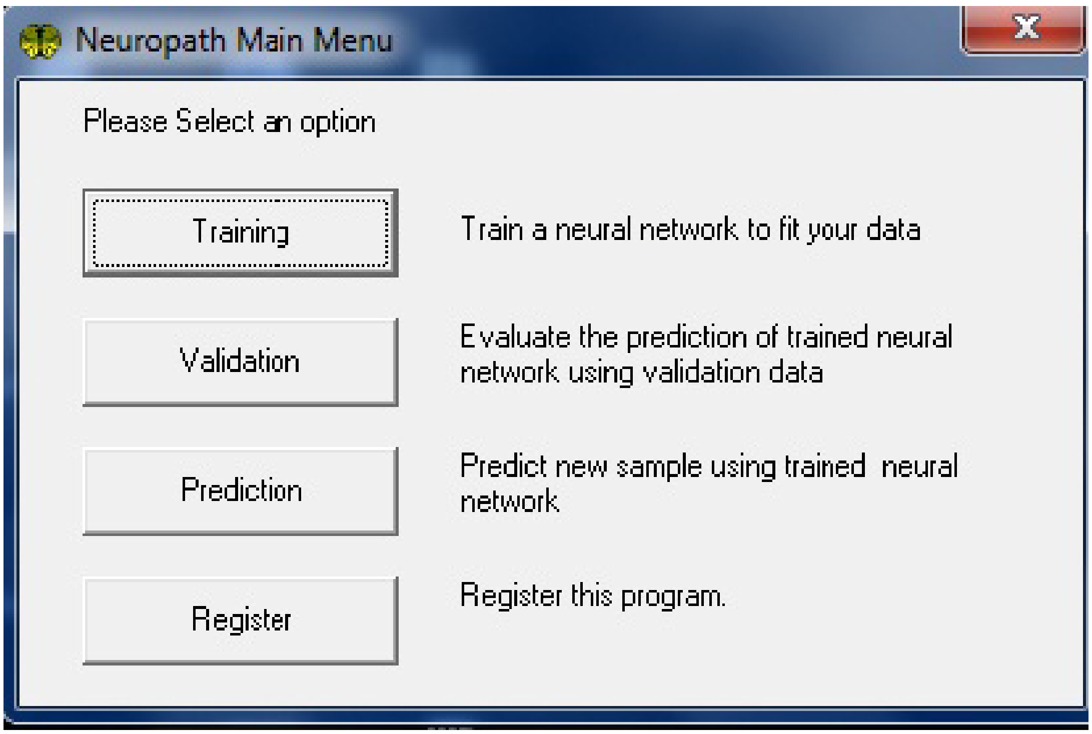


Fig 5.3 Start screen of *Neuropath* Software

For training the neural networks to fit the data "***Training***" button has to be clicked which will lead to the second screen of the program. Next we need to select a working directory. The data file should be located in the working directory. The data file is Selected to train by double click on the filename list. A preview of the file will appear on top of the dialog box. Enter the appropriate no. of inputs and outputs is entered as shown in figure 5.4.No. columns in the data file should equal to: No. Inputs + outputs + 1 (id column)

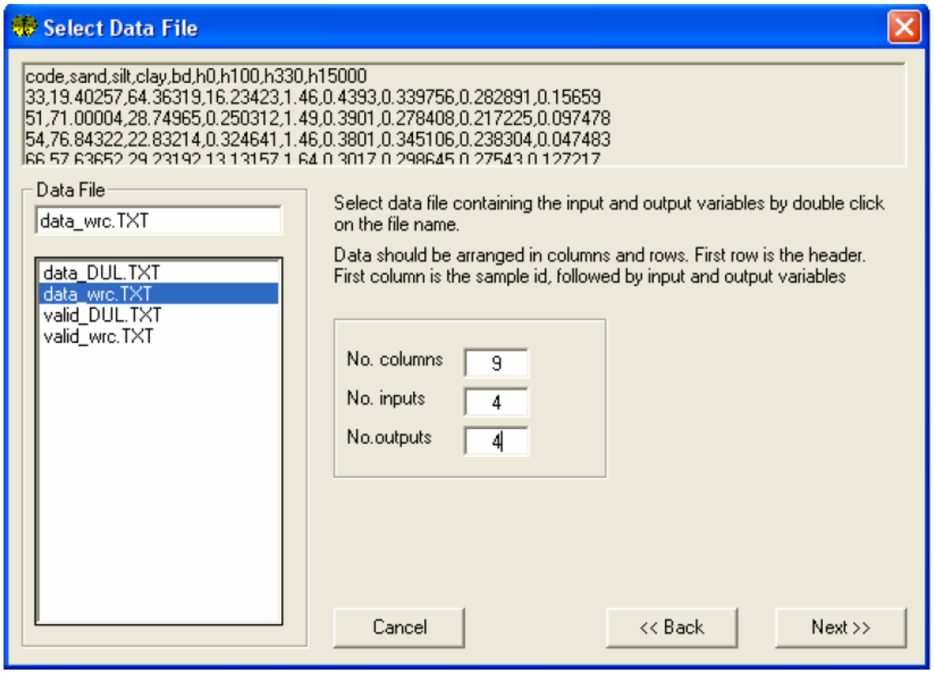


Fig 5.4 Selecting the data file

Neuropath software will produce several output files. All the outputs will be written in the working directory as shown in figure 5.5.

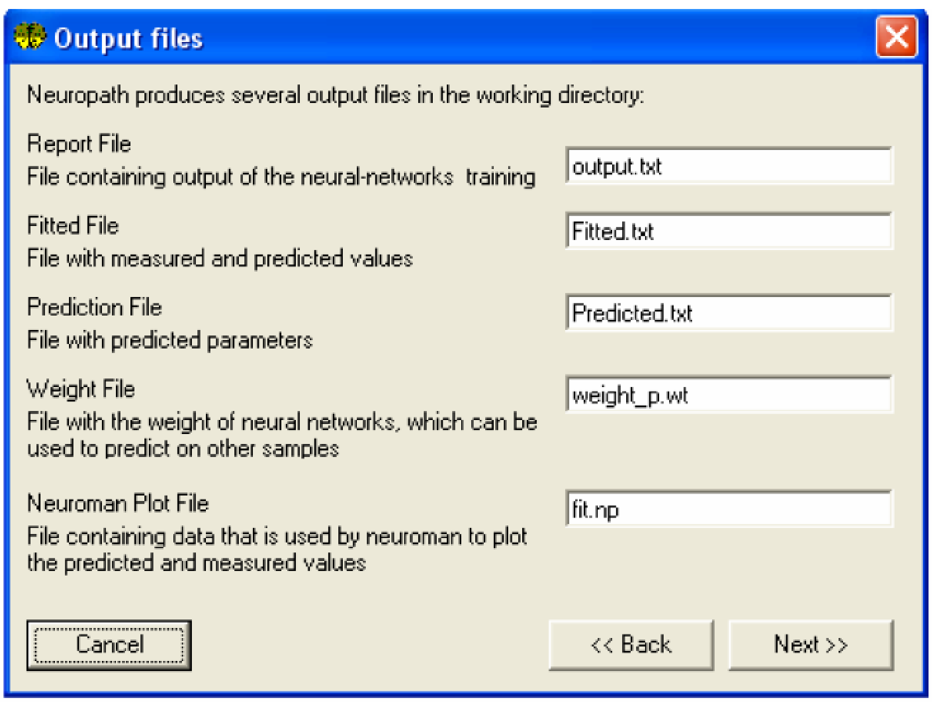


Fig 5.5 Output files generated by *Neuropath* Software

The output files produced are:

Output.txt
Shows the general information and also the objective function, AIC, and RMSE for each output at each bootstrap from the training data.

Out_of_bag.txt
Shows the objective function, AIC, and RMSE for each output at each bootstrap of the data not used in training. In bagging, predictors are constructed using bootstrap samples from the training set. Each bootstrap sample leaves out about 37% of the examples. These left-out data can be used to assess accuracy of the estimates Breiman (1997).

Fitted.txt
Shows the measured and fitted values. The first columns are the measured values, followed by the predicted starting with label “m_”, then followed by the standard deviations of prediction with prefix “s_”, and ME (mean error) & RMSE (Root Mean Squared Error).

5.2.2 Defining the neural network parameters

No. of hidden units controls the complexity of the network. Generally is better to have too many hidden units than too few. With too few, the network would not be able to capture the flexibility of the data.

Neuraopath software uses bootstrap bagging (bootstrap aggregating) predictors. Bagging predictors is a method for generating multiple versions of predictor and using these to get an aggregated predictor. It runs the neural network several times as defined by number of bootstrap using randomly perturbed versions of the data.

The final (bagged) estimates averages the prediction of different network. This in some way can alleviate the problem of falling into single local minima, as objective function can posses many local minima.

The optimal number of hidden units is found by trial and error and set to 5 in this study. While a conservative number of 50 is selected for defining the number of bootstrap for prediction. Number of iterations is the maximum number of iterations that nonlinear least squares can attempt to decrease the objective function.

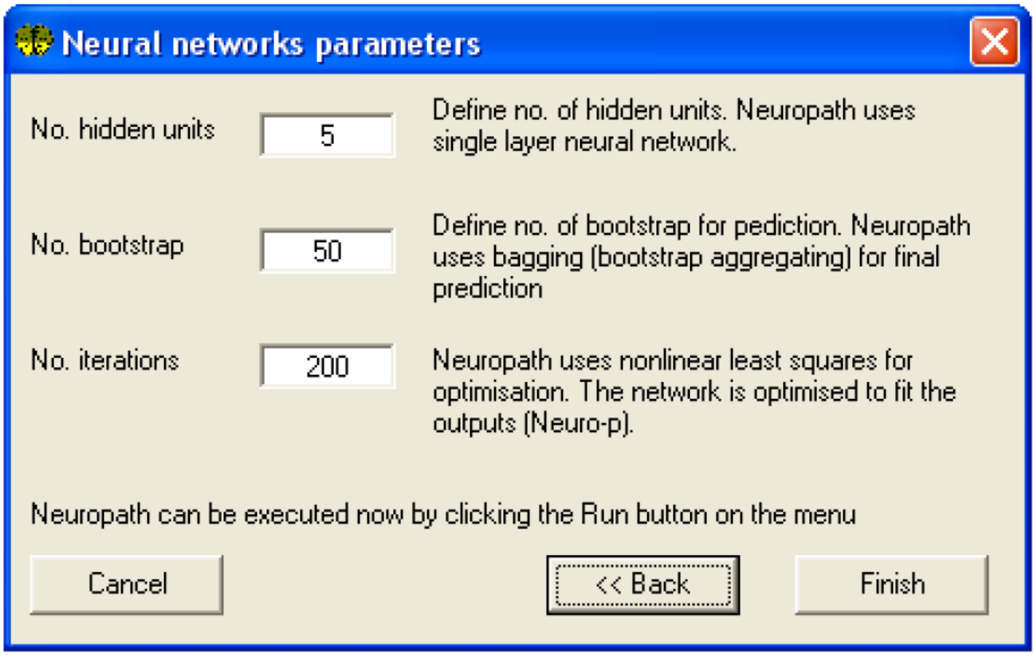


Fig 5.6 Selecting the number of neurons in hidden layer and number of bootstraps

When all the required inputs/parameters have been entered properly, the program will display the main dialog box with all the entered parameter values (Figure 5.7).

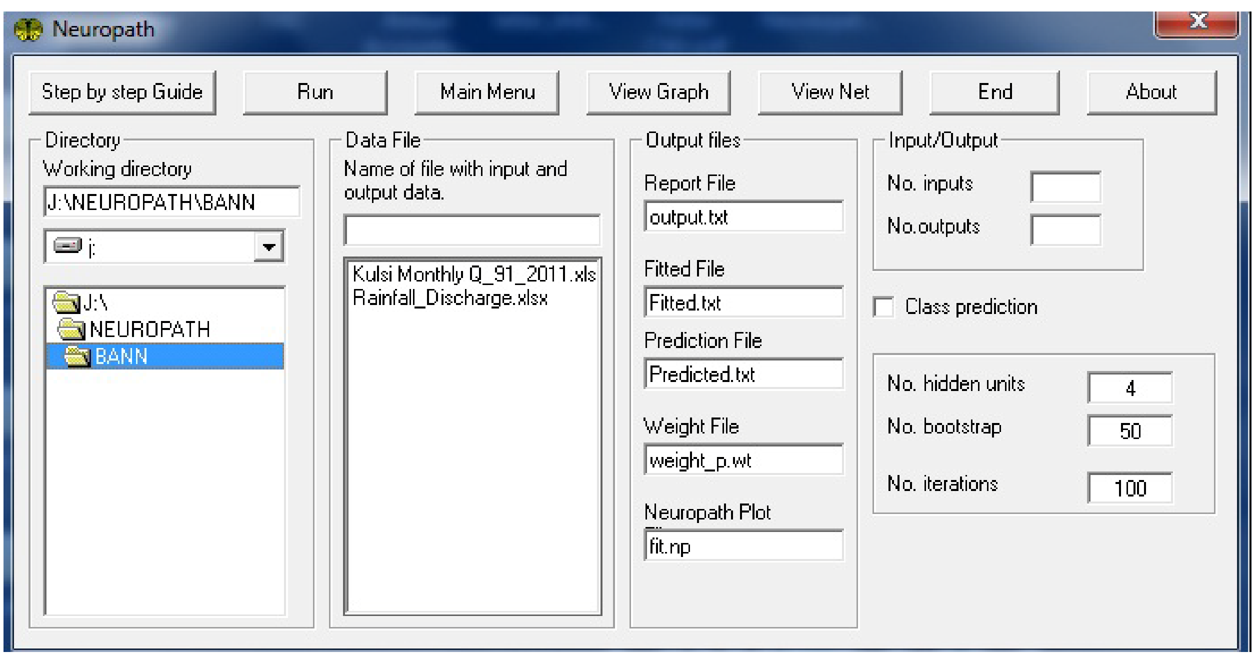


Figure 5.7 Main Dialog box with all entered parameters.

The program can be executed by clicking the “Run” button. The Fortran program will run and at the end a message box will show “Program terminated with exit code 0. Exit windows?” This means the program is successfully executed; click yes to close the Fortran program (figure 5.8).

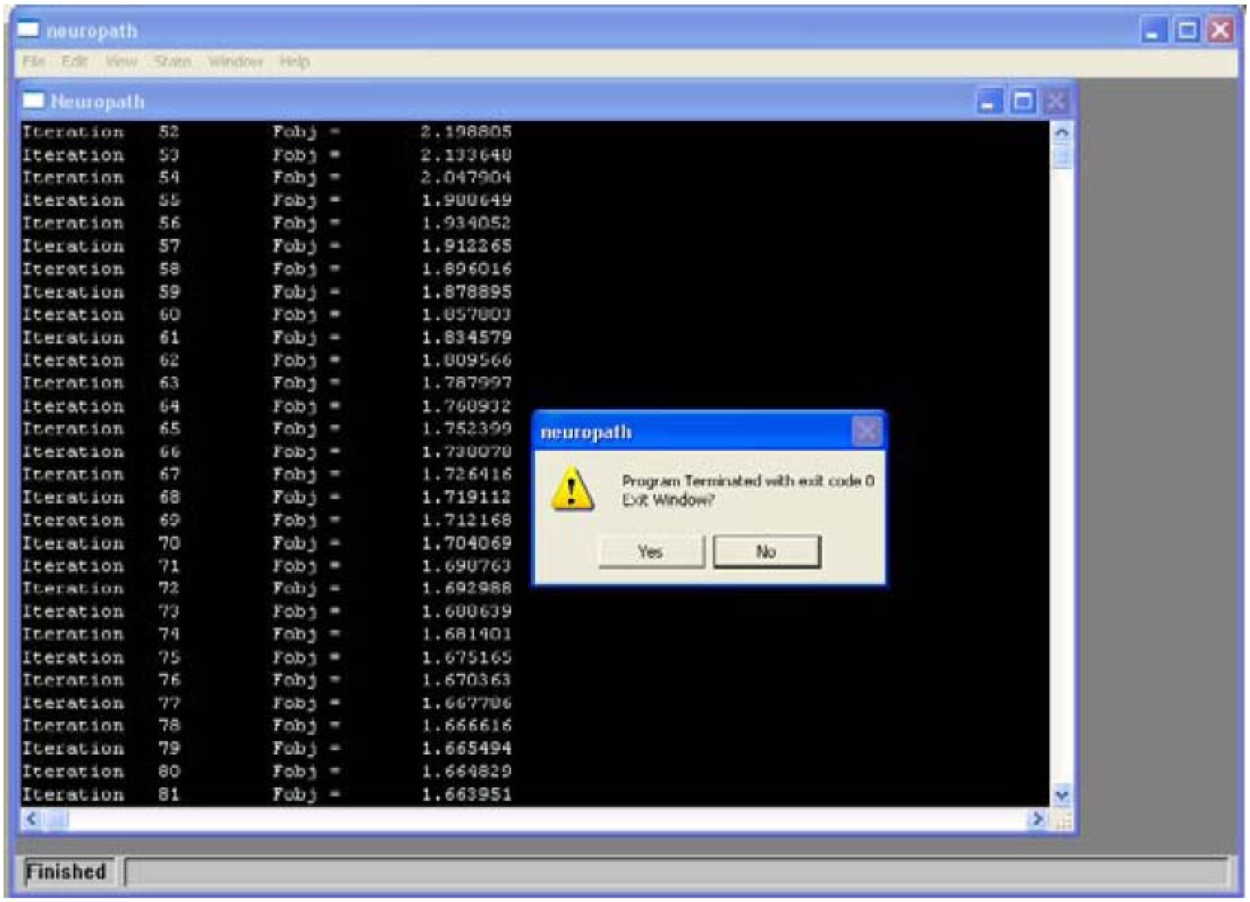


Figure 5.8 Execution of *Neuropath* Software

5.3 Performance Indices

The outputs were destandardized to generate the predicted values of discharge values for comparison with the observed values. Performance of the neural networks was evaluated by coefficient of determination (R^2) for the validation datasets. The uncertainty in neural networks was quantified by bootstrap estimates using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE).

6.0 RESULTS AND DISCUSSION

6.1 Data Pre-processing

Historical monthly rainfall data for 21 years (1990-2010) were procured by the Centre from Indian Meteorological Department (IMD), Guwahati, Assam. Concurrent data from five weather stations i.e. North Lakhimpur (S1), Mohanbari (S2), Guwahati (S3), Tezpur (S4) and Silchar (S5) from Assam and one weather station i.e. Shillong (S6) from Meghalaya were used in the study.

Shortest Straight line distance (km) between the weather stations is given in Table. 6.1. CWC maintains one Level II GDS Station at KulsiBazzar, the location of which is also shown in figure 6.1. The daily data of discharge and water level and sediment yield is available since 1972 with few years missing data.

Table 6.1 Shortest Straight Line Distance between Stations

Distance (km)	S1	S2	S3	S4	S5	S6
S1	0	110	338	187	394	390
S2		0	350	297	482	482
S3			0	153	236	83
S4				0	263	193
S5					0	166
S6						0

The coordinates of weather stations were imported into the ArcGIS^{TM} Software in Geographic Projection System as Shown in Figure 6.1.

The objective of data pre-processing is to predict the rainfall distribution at Kulsi Bazzar where the GDS site is maintained in order to get concurrent rainfall and discharge values. These values would be used in BANN models for simulation of discharge

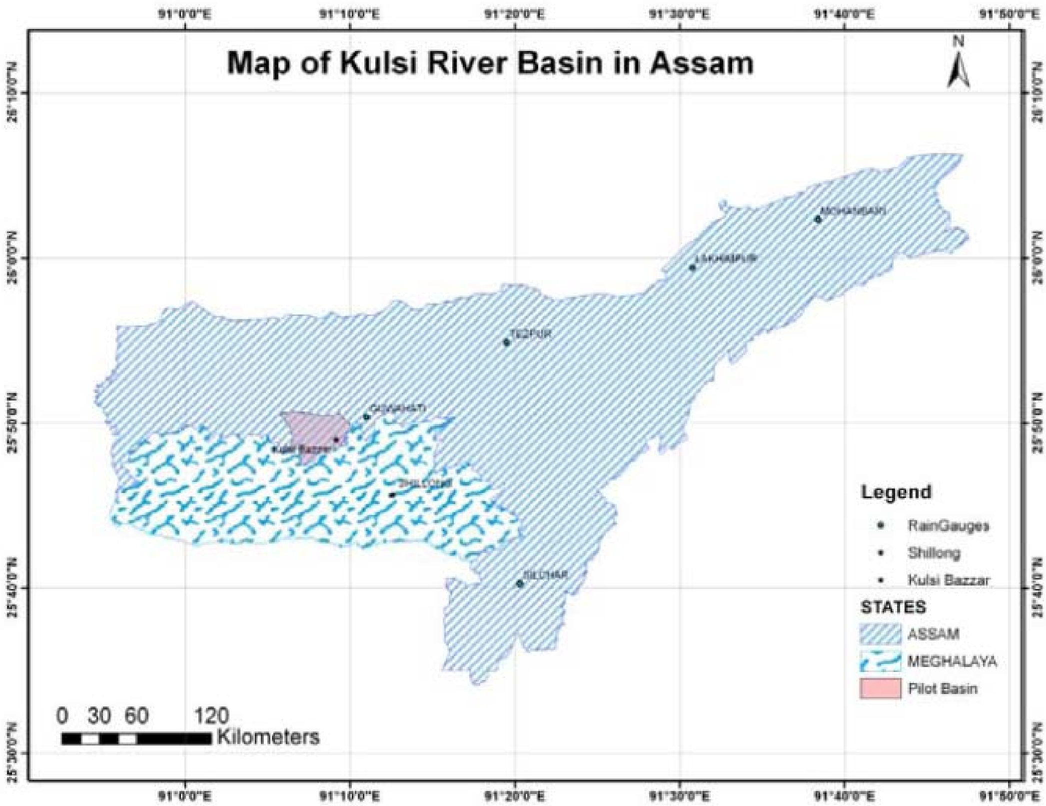


Figure 6.1 Location of Rain gauges in Assam and Meghalaya

6.2 Basic Statistics

The mean monthly (mm) rainfall averages from stations (S1 to S5) falling in Assam are given in Figure 6.2 The annual rainfall of the weather stations is shown in Fig 6.3

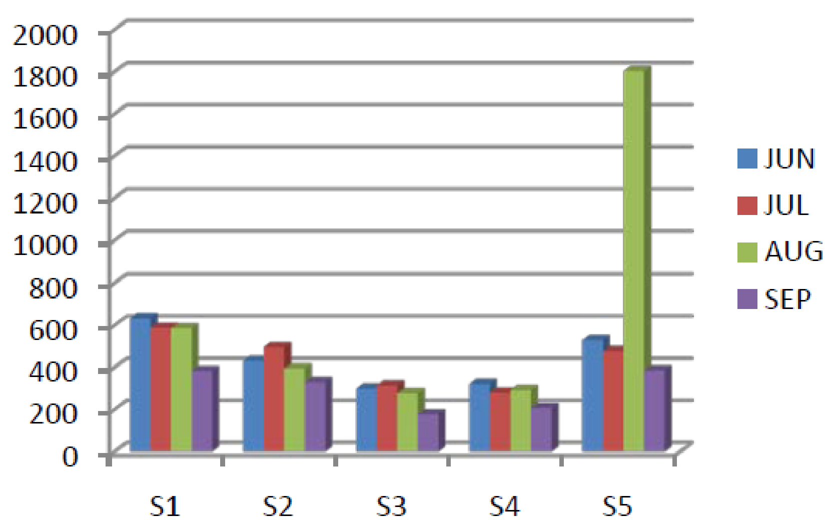


Fig 6.2 Monthly Rainfall Average Across Assam

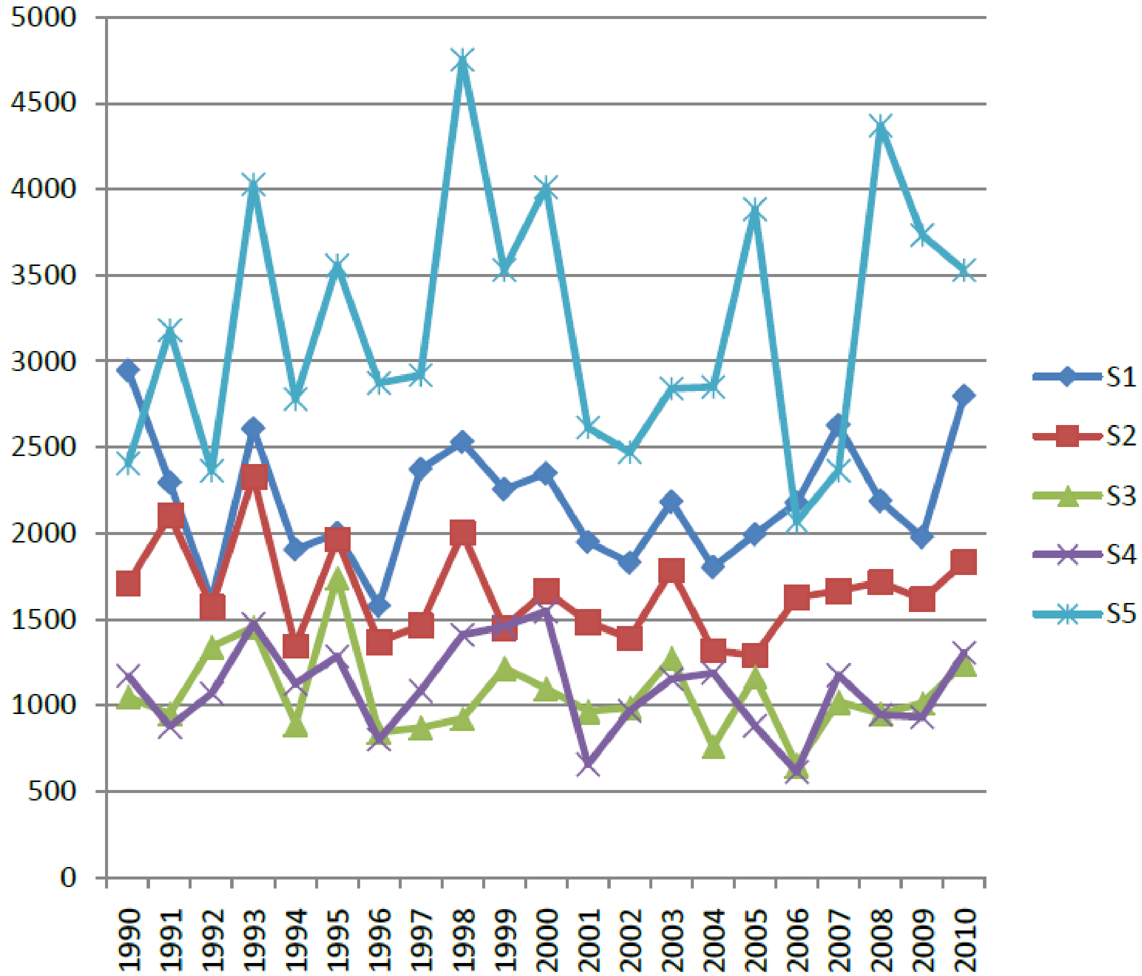


Fig. 6.3 Annual Rainfall of Selected Stations across Assam

The descriptive statistics of S1 to S6 stations falling in Assam and Meghalaya is given in Table 6.2.

The Pearson Correlation at 0.01 and 0.05 level are determined using SPSS^{TM} software given in table 4.3.

Table 6.2 Descriptive Statistics of Weather Stations

	N	Minimum (mm)	Maximum (mm)	Mean (mm)	Std. Deviation (mm)
S1 JUN	40	217	1021	606	188
S1 JUL	40	419	1148	659	172
S1 AUG	40	310	1062	548	169
S1 SEP	40	116	816	427	168
S2 JUN	40	148	913	406	152
S2 JUL	40	249	768	523	111
S2 AUG	40	139	752	414	145
S2 SEP	40	85	577	335	127
S3 JUN	40	95	550	306	115
S3 JUL	40	175	646	338	117
S3 AUG	40	65	803	257	127
S3 SEP	40	28	372	186	82
S4 JUN	40	80	606	302	123
S4 JUL	40	105	638	338	110
S4 AUG	40	87	518	292	114
S4 SEP	40	85	478	229	103
S5 JUN	40	263	4705	2445	880
S5 JUL	40	1341	6667	3050	1169
S5 AUG	40	623	4659	1786	833
S5 SEP	40	285	4660	1160	808
S6 JUN	40	87	769	427	179
S6 JUL	40	119	1078	460	262
S6 AUG	40	87	682	297	123
S6 SEP	40	123	920	289	143

Table 6.3 Person Correlation Coefficients

Correlations for Month of June

		S1 JUN	S2 JUN	S3 JUN	S4 JUN	S6 JUN	S5 JUN
S1 JUN	Pearson Correlation	1	**.366***	-.019	**.411****	-.159	.278
	Sig. (2-tailed)		.020	.910	.008	.328	.199
	N	40	40	40	40	40	23
S2 JUN	Pearson Correlation	**.366***	1	-.112	**.352***	-.120	**.454***
	Sig. (2-tailed)	.020		.493	.026	.461	.029
	N	40	40	40	40	40	23
S3 JUN	Pearson Correlation	-.019	-.112	1	.260	**.465****	.002
	Sig. (2-tailed)	.910	.493		.106	.003	.991
	N	40	40	40	40	40	23
S4 JUN	Pearson Correlation	**.411****	**.352***	.260	1	.083	.412
	Sig. (2-tailed)	.008	.026	.106		.609	.050
	N	40	40	40	40	40	23
S6 JUN	Pearson Correlation	-.159	-.120	**.465****	.083	1	-.221
	Sig. (2-tailed)	.328	.461	.003	.609		.311
	N	40	40	40	40	40	23
S5 JUN	Pearson Correlation	.278	**.454***	.002	.412	-.221	1
	Sig. (2-tailed)	.199	.029	.991	.050	.311	
	N	23	23	23	23	23	23

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Correlations for Month of July

		S1 JUL	S2 JUL	S3 JUL	S4 JUL	S6 JUL	S5 JUL
S1 JUL	Pearson Correlation	1	**.436****	**.441****	**.458****	**.470****	.077
	Sig. (2-tailed)		.005	.004	.003	.002	.728
	N	40	40	40	40	40	23
S2 JUL	Pearson Correlation	**.436****	1	.200	**.340***	.251	-.129
	Sig. (2-tailed)	.005		.215	.032	.119	.558
	N	40	40	40	40	40	23
S3 JUL	Pearson Correlation	**.441****	.200	1	**.533****	**.383***	.033
	Sig. (2-tailed)	.004	.215		.000	.015	.882
	N	40	40	40	40	40	23
S4 JUL	Pearson Correlation	**.458****	**.340***	**.533****	1	**.411****	**.417***
	Sig. (2-tailed)	.003	.032	.000		.008	.048
	N	40	40	40	40	40	23
S6 JUL	Pearson Correlation	**.470****	.251	**.383***	**.411****	1	**.532****
	Sig. (2-tailed)	.002	.119	.015	.008		.009
	N	40	40	40	40	40	23
S5 JUL	Pearson Correlation	.077	-.129	.033	**.417***	**.532****	1
	Sig. (2-tailed)	.728	.558	.882	.048	.009	
	N	23	23	23	23	23	23

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Correlations for Month of August

		S1 AUG	S2 AUG	S3 AUG	S4 AUG	S6 AUG	S5 AUG
S1 AUG	Pearson Correlation	1	.274	.350*	.204	.426**	.139
	Sig. (2-tailed)		.087	.027	.208	.006	.394
	N	40	40	40	40	40	40
S2 AUG	Pearson Correlation	.274	1	.224	.396*	.565**	.229
	Sig. (2-tailed)	.087		.164	.011	.000	.155
	N	40	40	40	40	40	40
S3 AUG	Pearson Correlation	.350*	.224	1	.369*	.477**	.275
	Sig. (2-tailed)	.027	.164		.019	.002	.086
	N	40	40	40	40	40	40
S4 AUG	Pearson Correlation	.204	.396*	.369*	1	.539**	.334*
	Sig. (2-tailed)	.208	.011	.019		.000	.035
	N	40	40	40	40	40	40
S6 AUG	Pearson Correlation	.426**	.565**	.477**	.539**	1	.629**
	Sig. (2-tailed)	.006	.000	.002	.000		.000
	N	40	40	40	40	40	40
S5 AUG	Pearson Correlation	.139	.229	.275	.334*	.629**	1
	Sig. (2-tailed)	.394	.155	.086	.035	.000	
	N	40	40	40	40	40	40
	N	23	23	23	23	23	23

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Correlations for Month of September

		S1 SEP	S2 SEP	S3 SEP	S4 SEP	S6 SEP	S5 SEP
S1 SEP	Pearson Correlation	1	.531**	.456**	.472**	.498**	.560**
	Sig. (2-tailed)		.000	.003	.002	.001	.005
	N	40	40	40	40	40	23
S2 SEP	Pearson Correlation	.531**	1	.576**	.267	.406**	.508*
	Sig. (2-tailed)	.000		.000	.096	.009	.013
	N	40	40	40	40	40	23
S3 SEP	Pearson Correlation	.456**	.576**	1	.408**	.295	.199
	Sig. (2-tailed)	.003	.000		.009	.064	.362
	N	40	40	40	40	40	23
S4 SEP	Pearson Correlation	.472**	.267	.408**	1	.391*	.547**
	Sig. (2-tailed)	.002	.096	.009		.013	.007
	N	40	40	40	40	40	23
S6 SEP	Pearson Correlation	.498**	.406**	.295	.391*	1	.233
	Sig. (2-tailed)	.001	.009	.064	.013		.284
	N	40	40	40	40	40	23
S5 SEP	Pearson Correlation	.560**	.508*	.199	.547**	.233	1
	Sig. (2-tailed)	.005	.013	.362	.007	.284	
	N	23	23	23	23	23	23

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

6.3 Processing of Hydrometeorological Data

The processing of precipitation data (S1 to S5) in Assam was carried out using Geostatistical Tool Box of ArcGIS software. The trend analysis in North-South and East West Directions for the months of June, July, August and September are shown in Figure 6.4.

The blue line represents the trend in N-S direction and the green line displays trend in E-W direction. The visual inspection of graphs clearly indicates trends in both N-S and E-W directions.

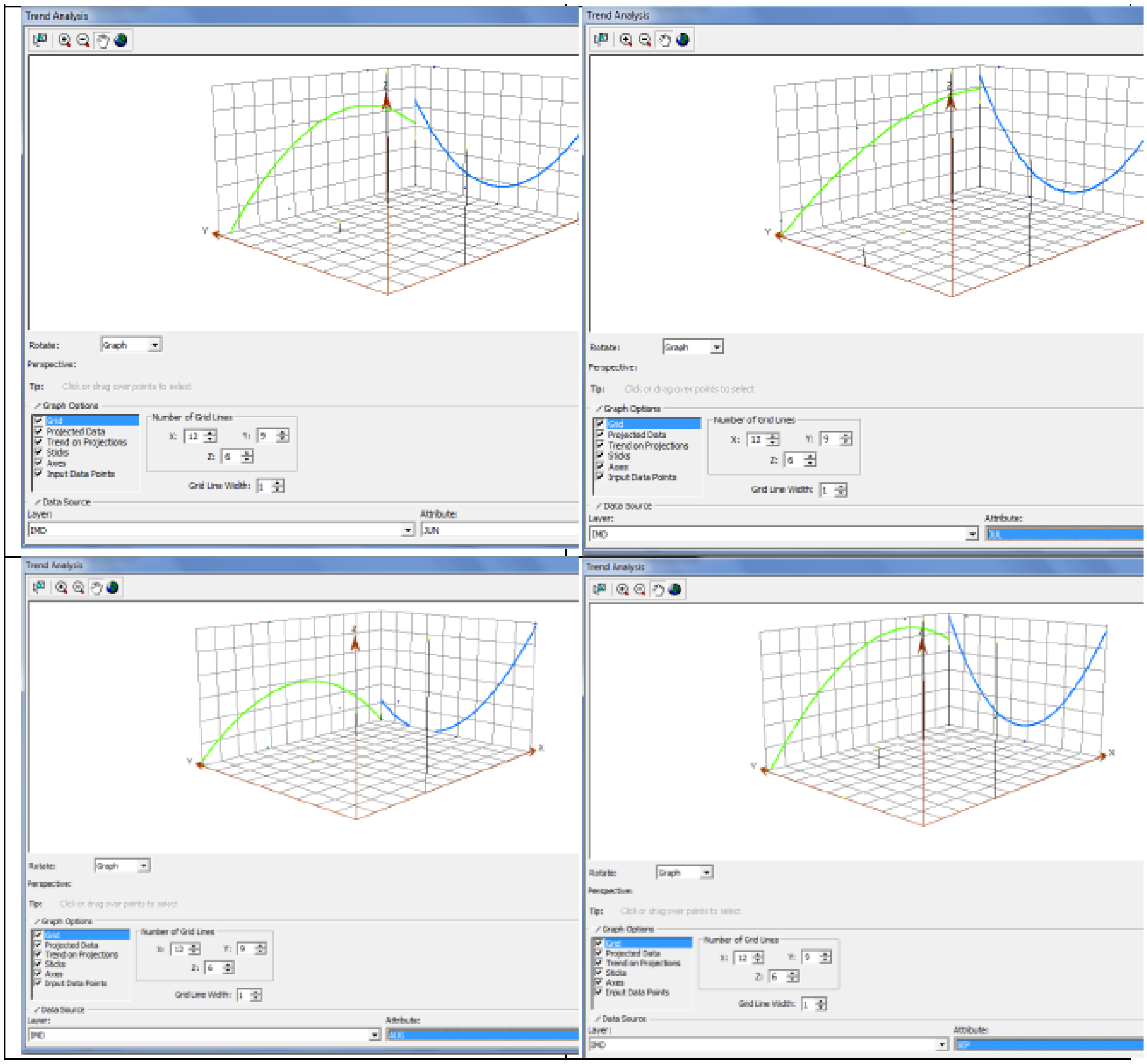
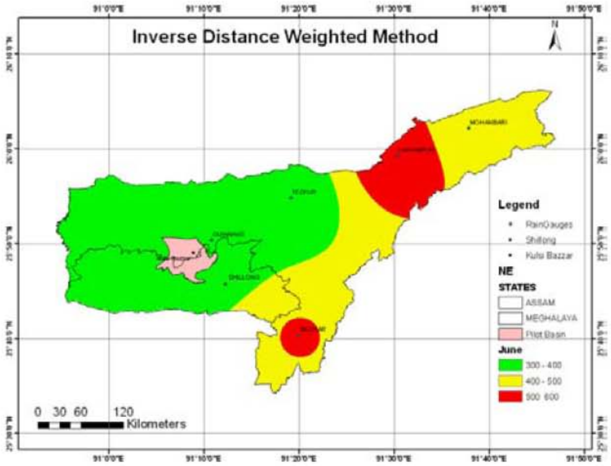


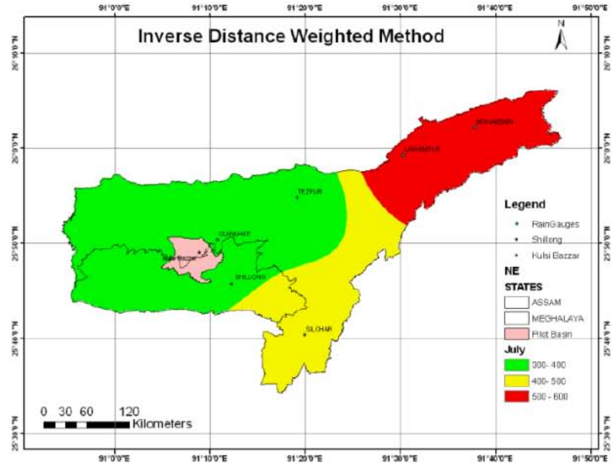
Fig 6.4 Trends in N – S and E – W Directions of Weather Stations (S1 to S5)

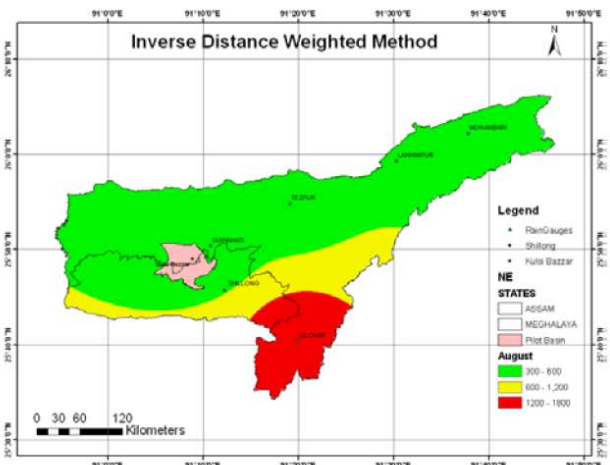
6.4 Spatial Analysis

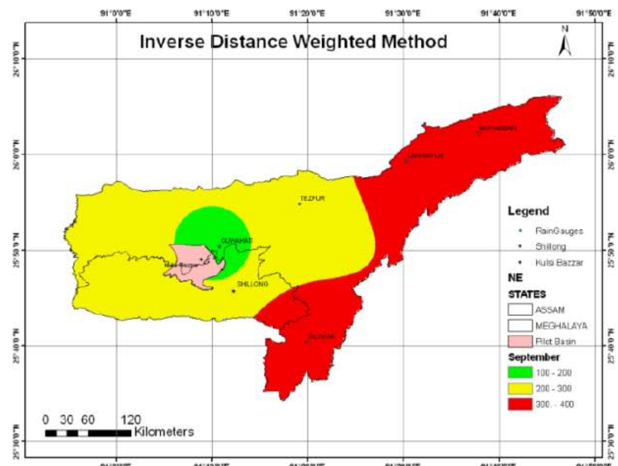
Spatial analysis of weather stations (S1 – S5) in Assam was carried out using Spatial Analyst Toolbox of ArcGIS™ software. Weather station (S6) in Meghalaya was excluded from the analysis in order to validate the model. Options for carrying out Inverse Distance Weighted model and Spline model are inbuilt in the toolbox.

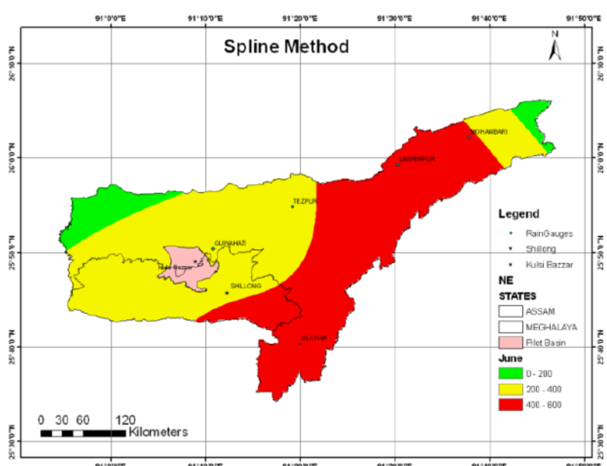
Krigging and Cokrigging methods of interpolations could not be executed in this study due limited number of sampling points. Figure 6.5 shows the spatial distribution of rainfall (mm) obtained based on the interpolation models. The sparse and un-uniform distribution of sampling points is one of the limitations of this study.

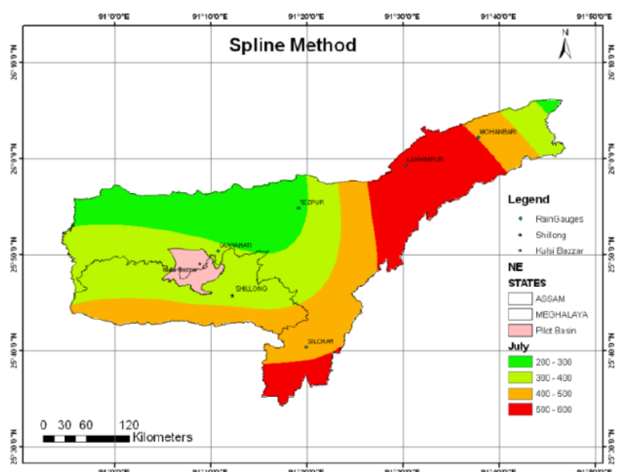

Inverse Distance Weighted Method
N
Legend
RainGauges
Shillong
Kulsi Bazzar
NE
STATES
ASSAM
MEGHALAYA
Pilot Basin
June
300 - 400
400 - 500
500 - 600
0 30 60 120
Kilometers


Inverse Distance Weighted Method
N
Legend
RainGauges
Shillong
Kulsi Bazzar
NE
STATES
ASSAM
MEGHALAYA
Pilot Basin
July
300 - 400
400 - 500
500 - 600
0 30 60 120
Kilometers


Inverse Distance Weighted Method
N
Legend
RainGauges
Shillong
Kulsi Bazzar
NE
STATES
ASSAM
MEGHALAYA
Pilot Basin
August
300 - 600
600 - 1,200
1200 - 1800
0 30 60 120
Kilometers


Inverse Distance Weighted Method
N
Legend
RainGauges
Shillong
Kulsi Bazzar
NE
STATES
ASSAM
MEGHALAYA
Pilot Basin
September
100 - 200
200 - 300
300 - 400
0 30 60 120
Kilometers


Spline Method
N
Legend
RainGauges
Shillong
Kulsi Bazzar
NE
STATES
ASSAM
MEGHALAYA
Pilot Basin
June
0 - 200
200 - 400
400 - 600
0 30 60 120
Kilometers


Spline Method
N
Legend
RainGauges
Shillong
Kulsi Bazzar
NE
STATES
ASSAM
MEGHALAYA
Pilot Basin
July
200 - 300
300 - 400
400 - 500
500 - 600
0 30 60 120
Kilometers

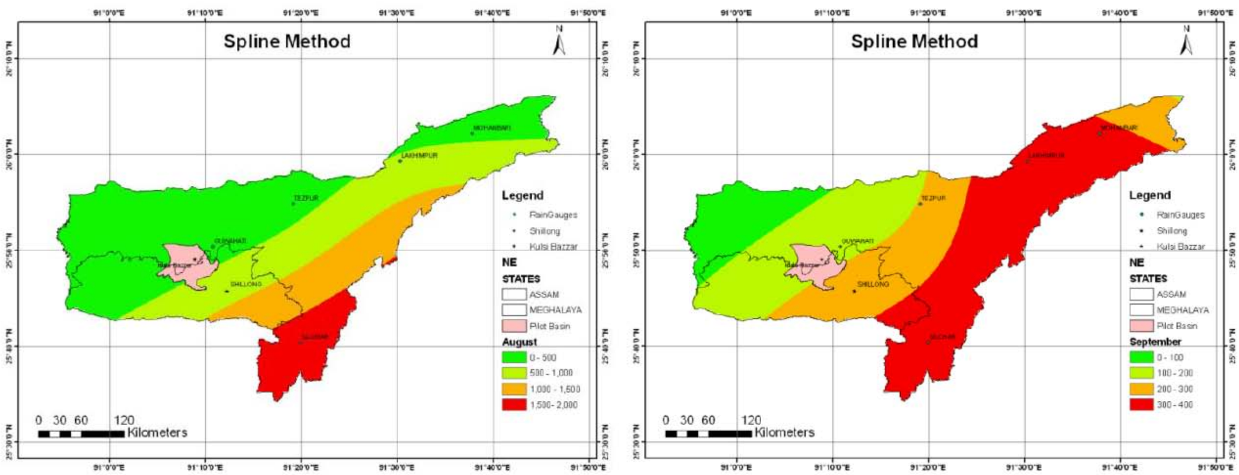


Figure 6.5 Spatial distribution of rainfall (mm) based on IDW and Spline Method

6.5 Performance evaluation for models

The grid values corresponding to weather station (S6) were obtained using the identify tool option in ArcGIS software. The rainfall values were identified and tabulated for the months of Jun - Sept. Trend lines were fitted between the estimated and measured monthly rainfall corresponding to Shillong weather station (S6) as shown in Figure 6.6

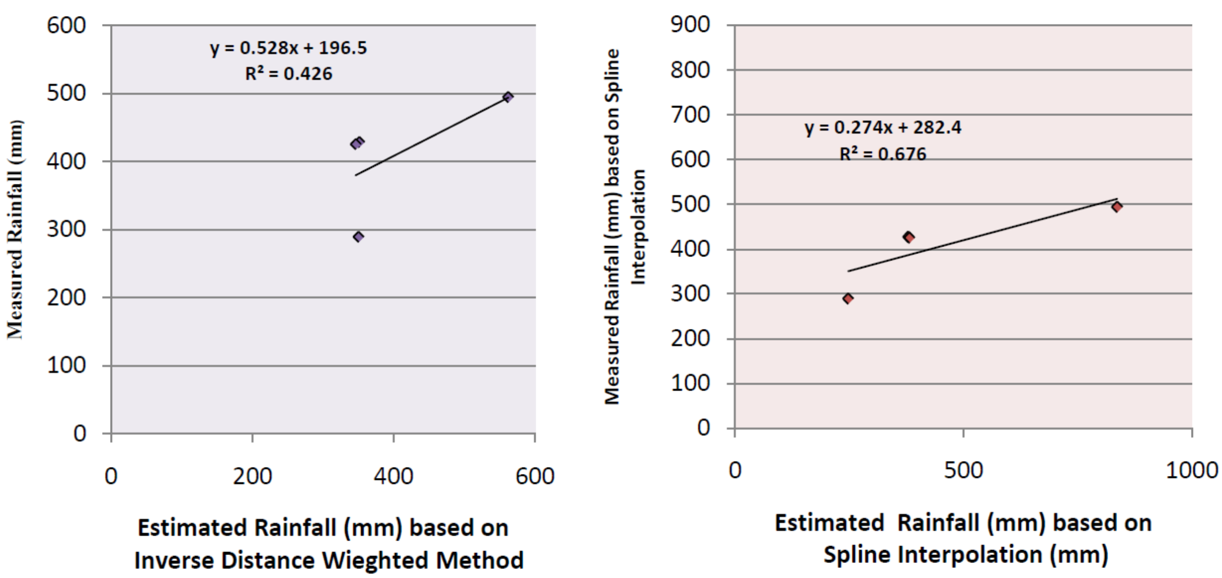


Fig 6.6 Estimated and Measured Rainfall (mm) based on IDW and Spline Interpolation

The inverse distance weighted method of interpolation produced ‘bull-eye’ type of pattern while the spline method of interpolation generated folded ‘rubber sheet” types of pattern. The performance of both the interpolation models were determined by selecting an independent station, (S6) corresponding to Shillong, Meghalaya.

The trend lines depicted Spline model (R^2 = 0.676) to perform better than the inverse distance method of interpolation(R^2 = 0.426). Selecting spline method of interpolation, the collinear values at KulsiBazzar (GDS II) monthly rainfall values for June to September were determined to be 308 mm, 339 mm, 289 mm and 176 mm respectively.

Similar procedure was repeated till the annual rainfall series for KulsiBazzar was obtained which was used as input in BANN model.

6.6 Setting up of BANN models

BANN models were developed using daily rainfall data from five weather stations. Training of BANN models was done for period of 2006 and validation was done for year 2007. 50 bootstrap replicates with 4 number of neurons in hidden layer was used for Training. The BANN models where developed following a hierarchical approach. BANN model 1 corresponded to only rainfall as inputs. Second hierarchy of BANN model 2 was developed with daily rainfall and rainfall lagged by one day as input. Similarly BANN model 3 was developed with inputs of daily rainfall, rainfall lagged by one day and rainfall lagged by two days. The process was repeated until the performance of BANN model was found to increase. The performance of BANN models was assessed by using coefficient of correlation (R^2), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). Table 6.4 shows the performance of models BANN1, BANN2 and BANN3. Further model development was stopped as the performance of the model was found to decrease after BANN2.

Table 6.4 Performance of BANN models of Validation Datasets

Model	**Inputs**	**R^2**	**RMSE**	**MAE**
BANN1	R_t	0.822	105.725	90.296
BANN2	***R_t, R_{t-1}***	***0.959***	***57.602***	***50.692***
BANN3	R_t, R_{t-1}, R_{t-2}	0.891	82.664	70.495

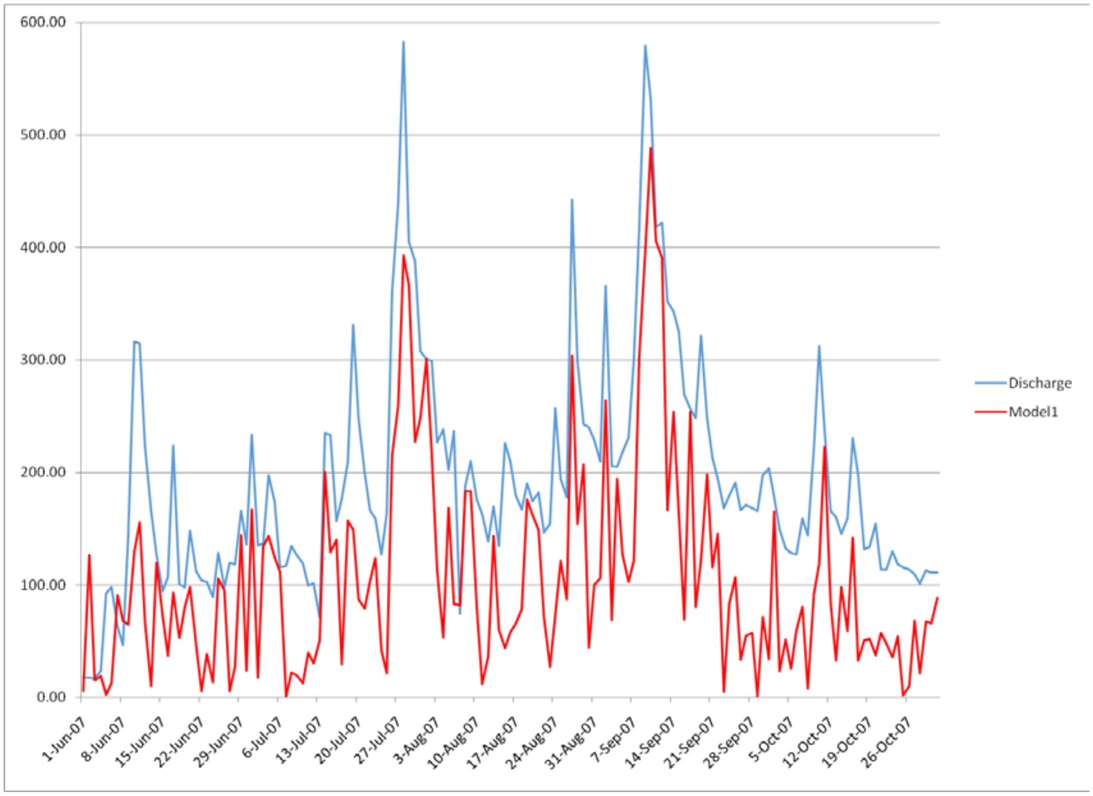


Fig 6.7 Performance of BANN1 with Observed Discharge

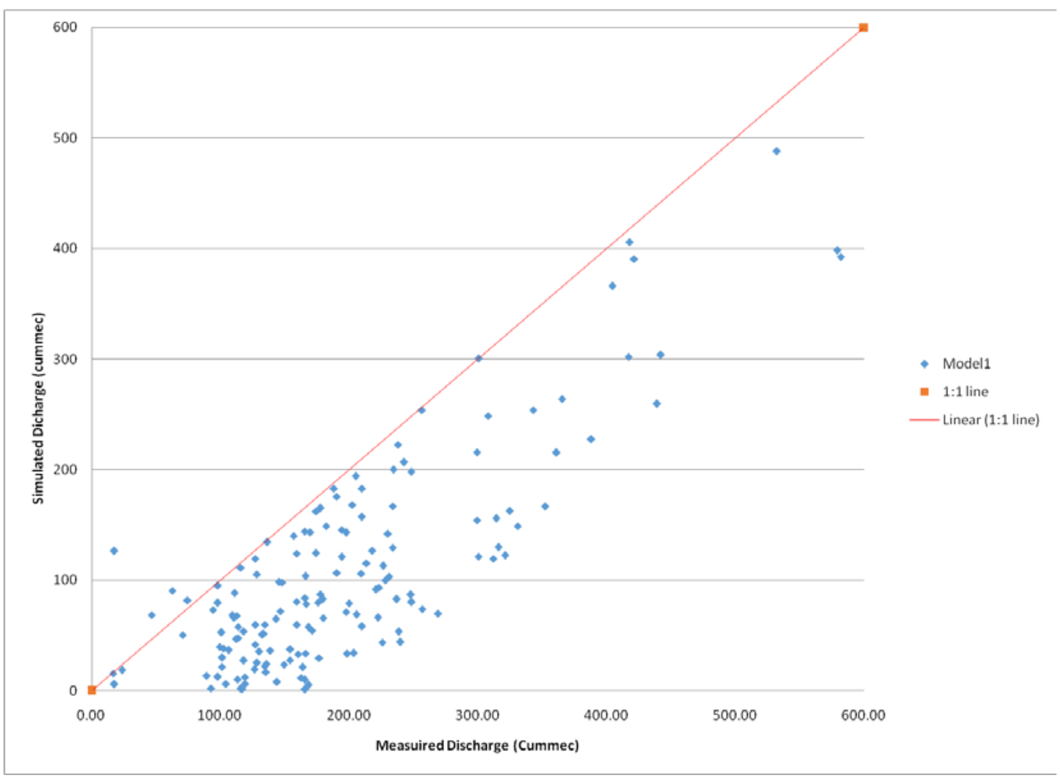


Fig 6.8 Scatter Plot of BANN1 with Observed Discharge

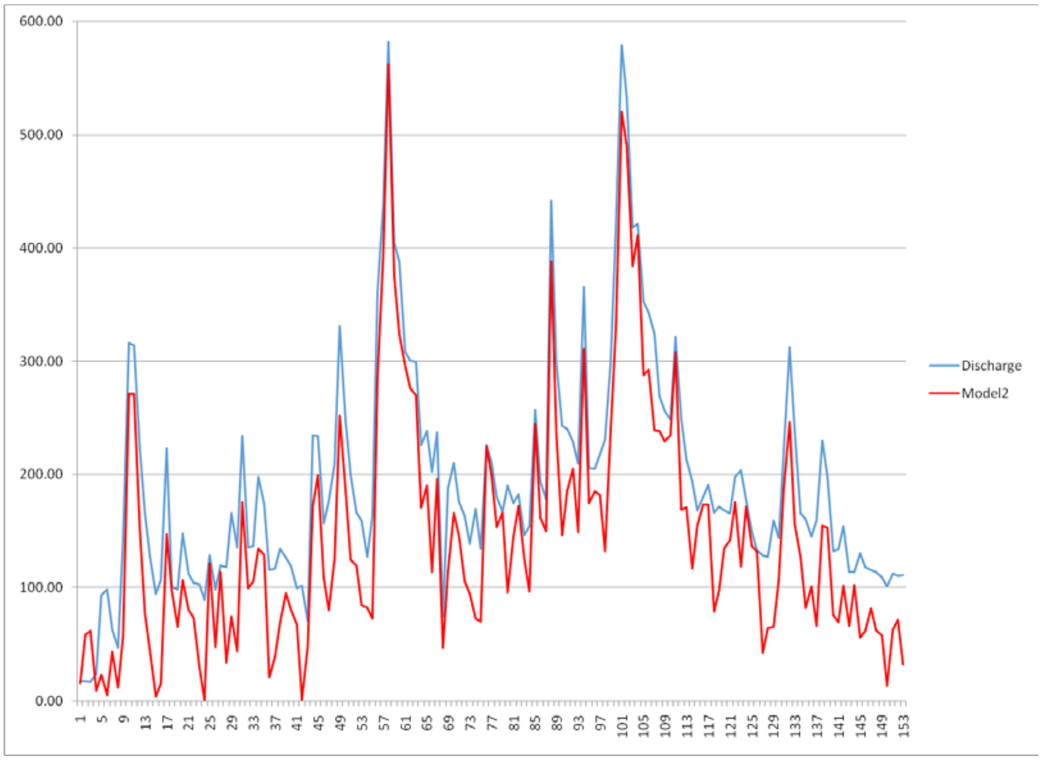


Fig 6.9 Performance of BANN2 with Observed Discharge

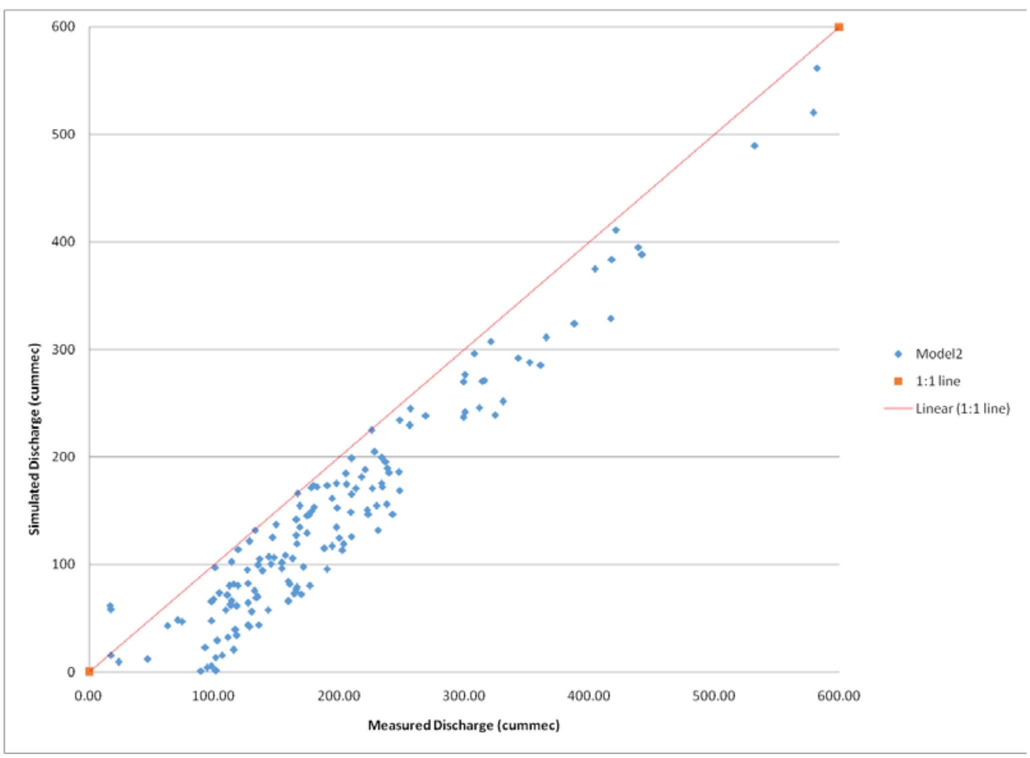


Fig 6.10 Scatter Plot of BANN2 with Observed Discharge

7.0 CONCLUSION

The following conclusions were drawn from the study:

1. BANN models were found to replicate the rise and fall of observed discharge with considerable accuracy. These results show the applicability of BANN models in forecasting floods when limited data is available.

2. Increasing the number of inputs in the model does not necessarily signify increase in the performance of BANN models. This trend was observed when BANN3 with three inputs showed lesser performance than BANN2 model with only two inputs.

3. All the BANN models were found to under predict the discharge values as observed from the scatterplots.

4. Best performance was observed for model with one day lagged rainfall values as input (BANN2) which shows the significance of adding lagged rainfall as input in the BANN model and finally

5. RMSE and MAE quantified the uncertainty in model predictions. These values indicate the range by in which the Bootstrapped ANN models simulations vary.

8.0 REFERENCES

Akhtar, M. K., Corzo, G. A., Andel, S. J. van and Jonoski, A. 2009. River flow forecasting with artificial neural networks using satellite observed precipitation pre-processed with flow length and travel time information: case study of the Ganges river basin. Hydrol. Earth Syst. Sci., 13:1607-1618.

Anctil, F., Perrin, C. and Andréassian, V. 2004. Impact of the length of observed records on the performance of NN and of conceptual parsimonious rainfall-runoff forecasting models. Environ. Modell. Softw., 19(4):357–368.

Archer, G. and Giovannoni, J. M. 1998, Statistical analysis with bootstrap diagnostics of atmospheric pollutants predicted in the APSIS experiment. Water, Air, and Soil Pollution., 106, 43-81.

A S C E Task Committee (2000a), Artificial Neural Networks in Hydrology. 1; Preliminary Concepts, Journal o f Hydrology Engineering, V o l . 5, No.2

A S C E Task Committee (2000b), Artificial Neural Networks in Hydrology. 11: Hydrologic Applications, Journal o f Hydrology Engineering, V o l . 5, No.2

Bailer, A. J. and Oris, J. T. 1997. Estimating inhibition concentrations for different response scales using generalized linear models. Environ. Toxicol., Chem. 16,. 1554–1559.

Bates, B. C. and Townley, L. R. 1988. Nonlinear, discrete flood event models: 3. Analysis of prediction uncertainty. J. Hydrol., 99: 91-101.

Beven, K. 2006. A manifesto for the equifinality thesis. J. Hydrol., 320:18-36.

Bonafe, A., Galeati, G. and Sforna, M. 1994. Neural networks for daily mean flow forecasting. In: W. R. Blain and K. L. Katsifarakis (editors), Hydraulic Engineering Software, Computational Mechanics Publications, Southampton, U.K., 1:131-138.

Breiman., 1996. Bagging predictors. Machine Learning 26, 123-140.

Bruen, M. and Yang, J. 2005. Functional networks in real-time flood forecasting-a novel application. J. Hydrol., 28:899-909.

Box, G. E. P., and Jenkins, G. M. 1976. Revised ed. Time series analysis: forecasting and control. Holden-Day, San Francisco.

Cameron, D., Kneale, P. & See, L. (2002), A n evaluation of a traditional and a neural net modelling approach to flood forecasting for an upland catchment, Hydrological Process, 16, 1033-1046.

Campolo, M . , Andreussi, P. & Soldati,A., (1999), River flood forecasting with a neural network. Water Resources Research, 35(4), 1191-1197.

Castellano-Mendez, M. 2004. Modelling of the monthly and daily behaviour of the runoff of the Xallas river using Box-Jenkins and neural networks methods. J. Hydrol., 296:38-58.

Cıgızoglu, H. K. 2003. Incorporation of ARMA models into flow forecasting by artificial neural networks. Environmetrics, 14(4):417-427.

Coulibaly, P. and Evora, N. D. 2007. Comparison of neural network methods for infilling missing daily weather records. J. Hydrol., 341:27-41.

Cover, K. A. and Unny, T. E. 1986. Application of computer intensive statistics to parameter uncertainty in streamflow synthesis, Water Resources Bulletin, 22, 3, 495-507.

Corzo, G. and Solomatine, D. P. 2007a. Knowledge-based modularization and global optimization of artificial neural network models in hydrological forecasting. Neural Networks, 20:528-536.

Dawson, C.W. & Wilby, R. (1998), A n artificial neural network approach to rainfall-runoff modelling, Hydrology Sciences, Joumal, 43, 47-66.

Dawson, C.W. & Wilby, R . L . (2001), Hydrological modelling using artificial neural networks, Progress in Physical Geography, 25(1), 80-108.

Demirel, M. C., Venancio, A. and Kahya, E. 2009. Flow forecast by SWAT model and ANN in Pracana basin, Portugal. J. Hydrol., 40:467-473.

Di Stefano, C., Ferro, V. and Porto, P. 2000. Applying the bootstrap technique for studying soil redistribution by caesium-137 measurements at basin scale, Hydrological Sciences Journal, 45 (2) 171–184

Dixon, P. M. 2001. The bootstrap and the jackknife: describing the precision of ecological indices. S. M. Scheiner and J. Gurevitch, editors. Design and analysis of ecological experiments, 2nd edition. pp 267–288, Oxford University Press, Oxford, UK.

Dybowski, R. and Roberts, S. J. 2000. Confidence and prediction intervals for feed forward neural networks. In: Dybowski R, Gant V. (Eds), Clinical Applications of Artificial Neural Networks. Cambridge University Press.

Efron, B. 1979. Bootstrap Methods: Another look at the Jackknife. Ann. Statist., 7:1-26.

Efron, B. and Tibshirani, R. J. 1993. An Introduction to the Bootstrap. Chapman and Hall, London, U.K.

Fortin, V., Bernier, J. and Bobe´e, B. 1997. Simulation, Bayes, and bootstrap in statistical hydrology, Water Resources Research., 33(3), 439– 448.

Georgakakos, K. P. and Krzysztofowicz, R. 2001. Probabilistic and ensemble forecasting. J. Hydrol., 249:1.

Halff, A. H., Halff, H. M. and Azmoodeh, M. 1993. Predicting runoff from rainfall using neural networks. Proc., Eng. Hydrol., ASCE, New York, 760–765.

Hall, M.J. and Minns, A.W. (1993) Rainfall-mnoff as a problem in artificial intelligence: experience with a neural network, Proc 4 th Nat. Hydrol. Symp., Cardiff British Hydrological Society, London, pp. 5.51-5.57.

Hsieh, W. W. and Tang, B. 1998. Applying neural network models to prediction and data analysis in Meteorology and Oceanography. Bulletin of the American Meteorological Society 79, 1855–1870.

Hettiarachchi, P., Hall, M. J. and Minns, A. W. 2005. The extrapolation of artificial neural networks for the modelling of rainfall-runoff relationships. J. Hydroinf., 7(4): 291-296.

Hjelmfelt A. T. and Wang M. 1993. Artificial neural networks as unit hydrograph applications. Proc. Eng. Hydrol., ASCE, 754-759.

Hsu, K., Gupta, H. V. and Sorooshian, S. 1995. Artificial Neural Network modeling of the rainfall-runoff process. Water Resour. Res., 31(10):2517–2530.

Hu, T. S., Lam, K. C. and Ng, S. T. 2001. River flow time series prediction with a range-dependent neural network. Hydrol. Sci. J., 46(5):729-745.

Imrie, C.E., Dumcan, S. & Korre, A. (2000), River flow prediction using artificial neural networks: generalisation beyond the calibration range, Joumal of Hydrology, 233, 138-153.

Jain, A. and Kumar, S. 2009. Dissection of trained neural network hydrologic models for knowledge extraction. Water Resour. Res., 45:W07420.

Kentel, E. 2009. Estimation of river flow by artificial neural networks and identification of input vectors susceptible to producing unreliable flow estimates. J. Hydrol., 375:481-488.

Lall, U. and Sharma, A. 1996. A nearest neighbor bootstrap for resampling hydrologic time series. Water Resources Research., 32, 3, 679-693.

Lange, N . T . G . (1998), Advantages of unit hydrograph derivation by neural networks, Hydroinformatics'98, Babovic, V . & Larsen, L . C . (eds), Balkema, Rotterdam, 783-789.

Laio, F., Porporato, A., Revelli, R. and Ridolfi, L. 2003. A comparison of nonlinear flood forecasting methods. Water Resour. Res., 39(5):1129.

Maier, H. R. and Dandy, G. C. 1997. Determining inputs for neural network models of multivariate time series. Microcomput. Civil Eng., 12:353-368.

Maier, H.R. & Dandy, G.C. (2000), Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications, Environmental Modelling & Software, (15), 101-124.

Minns, A . W . & Hall, M.J. (1996), Artificial neural networks as rainfall-mnoff models, Hydrological Science, Joumal, 41, 399-417

Minns, A . W . & Hall, M . J . (2003), Rainfall-runoff modelling, ch. 6 o f Abrahart, R.J., See,L. and Kneale, P.E. (eds) Neural networks for hydrological modelling, Balkema, Rotterdam

Moss, M. E. and Tasker, G. D. 1991. An intercomparison of hydrological network-design technologies, Hydrological Sciences Journal, 36, 3, 209-221.

Mount, N. J. and Abrahart, R. J. 2011. Discussion of ‘‘River flow estimation from upstream flow records by artificial intelligence methods’’ by M. E. Turan and M. A. Yurdusev [J. Hydrol., 369 (2009) 71–77]. J. Hydrol., 396:193-196.

Mutlu, E., Chaubey, I., Hexmoor, H. and Bajwa, S. G. 2008. Comparison of artificial neural network models for hydrologic predictions at multiple gauging stations in an agricultural watershed. Hydrol. Process., 22(26):5097-5106.

Sharma, S. K. and Tiwari, K. N. 2009. Bootstrap based artificial neural network (BANN) analysis for hierarchical prediction of monthly runoff in Upper Damodar Valley Catchment. J. Hydrol., 374:209-222.

Sharma, A., Tarboton, D. G. and Lall, U. 1997. Streamflow simulation: A nonparametric approach, Water Resources Research, 33, 3, 291-308.

Smith, J. and Eli, R. N. 1995. Neural network models of rainfall-runoff process. J. Water Resour. Plan. Manage., 121(6):499-508.

Smith, S. J. 1997. Bootstrap confidence limits for groundfish trawl survey estimates of mean abundance. Canadian Journal of Fisheries. Aquat. Sci. 54, 616–630.

Srinivas, V. V. and Srinivasan, K. 2000. Post-blackening approach for modelling dependent annual streamflows, Journal of Hydrology, 230, 86-126.

Srinivas, V. V. and Srinivasan, K. 2001. Post-blackening approach for modelling periodic streamflows, Journal of Hydrology, 241, 221-269

Solow, A. R. 1989. Bootstrapping sparsely sampled spatial point patterns. Ecology 70:379–382

Stone, M. 1974. Cross-validatory choice and assessment of statistical predictions (with discussion). J. Roy. Statist. Soc.Ser. B, 36:111-147.

Tasker, G. D. 1999. Bootstrapping Periodic ARMA Model to Forecast Streamflow at Multiple Sites, Computer Science and Statistics, 31, 296-299.

Tasker, G. D. and Dunne, P. 1997. Bootstrap position analysis for forecasting low flow frequency, Journal of Water Resources Planning and Management, 123, 6, 359-367.

Tokar, A. S. and Johnson, P. A. 1999. Rainfall-runoff modeling using artificial neural networks, J. Hydrol. Eng., 4(3):232-239.

Tokar, A. S. and Markus, M. 2000. Precipitation-runoff modelling using artificial neural networks and conceptual models. J. Hydrol. Eng., 5(2):156-161.

Vogel, R. M. and Shallcross, A. L. 1996. The moving blocks bootstrap versus parametric time series models, Water Resources Research, 32, 6, 1875- 1882.

Woo, M. K. 1989. Confidence intervals of optimal risk based hydraulic design parameters, Canadian Water Resources Journal, 14, 10-16.

Wu, C. L., Chau, K. W. and Fan, C. 2010. Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques. J. Hydrol., 389:146-167.

Zealand, C . M . , Bum, D . H . & Simonovic, S.P. (1999), Short term streamflow forecasting using artificial neural networks, Joumal o f Hydrology, 214, 32-48

Zhu, M., Fujita M. and Hashimoto, N. 1994. Application of neural networks to runoff prediction. In: Hipel K. W. et al (Eds) Stochastic and statistical method in hydrology and environmental engineering, Vol. 3, Kluwer, Dordrecht, The Netherlands, 205-216.

Zucchini, W. and Adamson, P. T. 1989. Bootstrap confidence intervals for design storms from exceedance series, Hydrological Sciences Journal, 34, 41-48.