HYBRID WAVELET TRANSFORM – NEURAL NETWORK APPROACH FOR SHORT TERM AND LONG TERM TIME SERIES FLOW FORECASTING

Thesis

Submitted in Partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

By

KHANDEKAR SACHIN DADU

(Register No. : 110633AM11P01)

DEPARTMENT OF APPLIED MECHANICS AND HYDRAULICS
NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA,
SURATHKAL, MANGALORE– 575025

DECEMBER – 2013

D EC L A R A T I O N

by the Ph.D. Research Scholar

I hereby *declare* that the Research Thesis entitled "**HYBRID WAVELET TRANSFORM-NEURAL NETWORK APPROACH FOR SHORT TERM AND LONG TERM TIME SERIES FLOW FORECASTING**" which is being submitted to the **National Institute of Technology Karnataka, Surathkal** in partial fulfillment of the requirements for the award of the Degree of **Doctor of Philosophy** in **Civil Engineering** is a *bonafide report of the research work carried out by me*. The material contained in this Research Thesis has not been submitted to any University or Institution for the award of any degree.

KHANDEKAR SACHIN DADU

(Reg. No. 110633AM11P01)

Department of Applied Mechanics and Hydraulics,

NITK, Surathkal.

Place: NITK, Surathkal

Date:

C E R T I F I C A T E

This is to certify that the Research Thesis entitled "**HYBRID WAVELET TRANSFORM-NEURAL NETWORK APPROACH FOR SHORT TERM AND LONG TERM TIME SERIES FLOW FORECASTING"** submitted by **KHANDEKAR SACHIN DADU** (Register Number: **110633AM11P01**) as the record of the research work carried out by *him, is accepted as the Research Thesis submission* in partial fulfillment of the requirements for the award of degree of **Doctor of Philosophy.**

Dr. Paresh Chandra Deka
Research Guide
(Name and Signature with date and seal)

Prof. Subba Rao
Chairman – DRPC
(Signature with date and Seal)

DEDICATED

TO MY WIFE *VARSHA*

ACKNOWLEDGEMENT

It is my pleasure to express profound gratitude and indebtedness towards my research supervisor **Dr. Paresh Chandra Deka**, Associate Professor, Department of Applied Mechanics and Hydraulics, for his continued inspiration, motivation, support, discussions and great patience throughout this research, which made this study possible. It is a valuable experience to learn many aspects from him as a good teacher. I admire among his other qualities, kindness and balanced approach towards success and failure; his scientific foresight and excellent knowledge have been crucial to the accomplishment of this work; who managed nicely to spare valuable time for guidance, valuable suggestions and excellent supervision of my research work. I consider myself privileged for having had the opportunity to conduct research in the area of soft computing techniques under his able supervision.

I am greatly indebted to Reseach Progress Appraisal Committee members, Prof. A. Mahesha, Department of Applied Mechanics and Hydraulics, Dr. Darshak Diwedi, Chemistry Department, for their critical evaluation and constructive comments and valuable suggestions during the progress of the work helped me to improve the quality of work.

I also extend my heartfelt thanks to Prof. Subba Rao, Head, Department of Applied Mechanics and Hydraulics, and Chairman DRPC for his continuous support, encouragement, and timely help, also for providing me all the necessary departmental facilities during my research period.

I gratefully acknowledge Prof. S. G. Mayya, Prof. Lakshman Nandgiri, Prof. M. K. Nagraj, Department of Applied Mechanics and Hydraulics, for their continuous support, care, timely help and their good wishes during the course of my work.

I am also grateful to all other faculty members, Department of Applied Mechanics and Hydraulics, NITK, Surathkal, for helping me directly or indirectly during my stay and research work.

I take this opportunity to express thanks to my friends Mr. Rahul Karale, Mr. Dattatray Mahajan, Mr. Santosh Babar, Mr. Amit Patil, and Mr. Prahlada R. (Research Scholars) for rendering my stay in the NITK Campus more than wonderful.

I also acknowledge the help and support provided non teaching staff, Sri. Jagadish B Foreman, Sri Balakrishna, Sri. Ananda Devadiga, Sri. Gopalakrishna, Sri. Padmanabha Acharya, Mr. Harish Salians, Mr. Harish D and Mrs. Prathima Prakash for their support and help during the research work.

The inspiration and support given by the other fellow Research Scholars of the Department of Applied Mechanics and Hydraulics have also been much appreciated.

I gratefully acknowledge Prof. M. N. Navale, Founder President, Sinhgad Technical Education Society, Pune, Dr./Mrs Sunanda M. Navale, Founder Secretary, Sinhgad Technical Education Society, Pune, and Dr. Suresh N. Mali, Principal, Sinhgad Institute of Technology and Science, Narhe, Pune, for their continuous support, timely help and their good wishes during the course of my work.

Without the support, patience and encouragement from my wife ***VARSHA,*** my son Tanish, daughter Anvi and my father "Anna" I could never have been able to submit this work.

KHANDEKAR SACHIN DADU

Place: NITK, Surathkal

Date:

ABSTRACT

Accurate modeling of runoff is useful in urban and environmental planning, flood and water resources management. In this research, a hybrid model has been developed for Brahmaputra River flow forecasting based on wavelet and artificial neural network (ANN) methods. In this current study, discrete wavelet transform was linked to ANN naming Wavelet Artificial Neural Network (WANN) for flow forecasting. Ten year daily flow data from January 1990 to December 1999 of Pandu and Pancharatna stations on Brahmaputra River, which carries heavy flood in monsoon season in the North-East region of India, were used in the study. The observed flow data were decomposed (up to 7 level) to multiresolution time series via discrete wavelet transform using Daubechies wavelets of order ranging from 1 (db1) to 5 (db5). Then multiresolution time series data were fed as input to ANN to get the forecasted discharge values. Daily data were used to forecast flow values for lead times 2, 3, 4, 7 and 14 day, weekly data were used to forecast flow values for lead times 1 week and 2 week, and monthly data were used to forecast flow values for lead time 1 month. The root mean square error (RMSE), determination coefficient (R^2), mean absolute error (MAE), BIAS (B), and scatter index (SI) were adopted to evaluate the model's performance. It was found that for all lead times WANN model has given better and consistent results compared to conventional ANN model. It was mainly because of multiresolution time series used as inputs. Also it was found that, model efficiency increases with increase in wavelet order, giving best results for db5 mother wavelet for all lead times for both the stations. Also, there has been significant impact of decomposition level on WANN model efficiency as observed in the study.

Keywords: Wavelet transform, artificial neural network, streamflow, Daubechies wavelet, time series.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

SYMBOLS

Symbol	**Description**
ANN	Artificial Neural Network
FFBP	Feed-forward back Propagation
LM	Levenberg-Marquardt
MLP	Multilayer Perception
NN	Neural Network
CWT	Continuous Wavelet Transform
DWT	Discrete Wavelet Transform
db*i*	Daubechies Wavelet of order *i*
FT	Fourier transform
STFT	Short Time Fourier Transform
WANN	Wavelet – Artificial Neural Network
B	BIAS
MAE	Mean Absolute Error
MSE	Mean Square Error
RMSE	Root Mean Square Error
R^2	Coefficient of Correlation
SI	Scatter Index

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

It is no longer possible to consider streamflow and other hydrologic processes as stationary (Milly et al, 2008). Nearly all the methods developed for the planning, management and operation of water resources systems assume stationarity of hydrologic processes. Non-stationarity can result from myriad human influences ranging from agricultural and urban land use modifications, to climate change and modifications to water infrastructure. Non-stationarity reduces uncertainty (because it explains part of variability). Inappropriate claim of non-stationarity results in underestimation of variability, uncertainty and risk. Analysis of hydrologic time series data and study of their fluctuations are two of the key issues to have an insight into the data, to identify pattern, trend or correlations and to understand the past and current behaviour of the flow. In addition to these, identifying the distribution of the analysed data is very important and necessary for understanding the dynamics of watershed behaviour as well as to determine when particular events will occur.

Hydrologic time series data appears to be noisy as well as non-stationary. The noise characteristic indicates the unavailability of complete information from past behaviour of flow conditions to fully capture dependency between future and past flow. The information that is excluded in the forecasting model is considered as noise while the non-stationary characteristic indicates the distribution of hydrologic time series changing over time. Therefore, hydrologic time series forecasting is considered as one of the most challenging tasks of time series analysis (Banhatti and Deka, 2012).

Hydrological Time series usually have seasonal variations and long-term and short-term fluctuations, which may not be limited to the mean of the series but may also effect its overall variance structure. Typically, during the rainy season (south west monsoon period in India), such series are characterized by patterns like trends and localized abrupt changes. The variations observed in hydrological time series are due to various hydrological processes operating in different time scales such as intraday, diurnal, several days, seasonal, annual and long term.

Short term variations may be due to local meteorological factors which cause the time series to become noisy and show periodicities of different scales. There are many hydraulic variables that effect flow and time series of these variables may possess the characteristics pertaining to non-stationary, non-linear and volatile time series having many outliers. Intermediate term variations may be due to more regional meteorological and seasonal factors. On the other hand, long term variations could be due to slow changes in urbanization and land use change over a long term period which effects the flow rate in the long run. Thus time series may present trend and nonstationary characteristics. Annual and inter-annual variations may be due to the weather and climate factors such as monsoon cycles in the tropical countries. In fact, the variations due to weather and meteorological factors are larger than variation due to long term changes in flow rate. Thus, to investigate the effectiveness of perennial river flow behaviour, one may need to separate the variations due to various factors in the time series.

River flow forecasting hours, days, months or possibly longer in advance is required for effective operation of water resources systems so that water authorities can administer water reserves optimally for various water users such as hydropower generation, agricultural, domestic, etc.(Deka and Chandramoulli, 2005). Flow forecasting is also important from view point of flood management. Forecasting of hydrological time series can be done by using stochastic models like Auto regressive (AR), Auto regressive moving average (ARMA) and Auto regressive integrated moving average (ARIMA) etc. These models are basically time series models and have a limited ability to capture nonstationarities and nonlinearities.

Recently soft computing techniques such as artificial neural network (ANN), fuzzy logic (FL) and genetic algorithm (GA) have been gaining popularity since last decade due to their versatility in handling non linearity and tp some extent to handle non stationary nature. Soft computing techniques offer effective approach for handling large amount of dynamic, non-linear and noisy data, especially when the underlying physical relationships are not fully understood (Nourani et al., 2011).

1.2 ARTIFICIAL NEURAL NETWORK MODEL

ANN is a mathematical model which mimics the function of human brain. It has the ability to identify the relationship from given patterns and solve large scale complex problems such as non-linear modeling pattern recognition, classification, association and control. Recently neural network models are successfully applied in rainfall-runoff modeling, runoff forecasting, evaporation estimation, precipitation forecasting, water quality modeling, ground water level forecasting, significant wave height forecasting and many others. Chandramouli and Deka (2005) developed decision support model using artificial neural network for optimal operation of a reservoir in South India. Tayfur and Singh (2006) predicted event based rainfall-runoff using artificial neural network and fuzzy logic models and compared the test results with kinematic wave approximation. The ASCE Task Committee (2000 II) reviews hydrologic applications of ANN.

ANN is suitable for handling large amount of dynamic, noisy and non-linear data, especially for partially understood underlying physical processes. This makes them effective to time series modelling problems of data-driven nature (Nourani et al., 2009 a). In spite of suitable flexibility of ANN, it may not be able to cope with non-stationary data if pre-processing of the input and output data is not performed (Cannas et al., 2006). As nonstationary signals are frequently encountered in a variety of engineering fields such as water resources and earthquake, hybridization of ANN with other techniques may provide effective modelling. ANN as black-box models, with two advantages of low quantitative data demands and simple formulation (Cheng et al., 2008), are used widely for the task. However, they have many drawbacks due to their black-box properties and also they are not suitable when the data is non-stationary.

For overcoming these drawbacks, many methods are often combined with black box models to forecast hydrologic time series. Present studies and practical applications indicate better performance of hybrid models rather than single model. Among the former, combination of wavelet analysis with black-box models has become a prevalent approach to conduct hydrologic time series forecasting.

1.3 WAVELET TRANSFORM

With regards to spectral analysis of time series, the introduction of wavelet functions has triggered new light into the analysis of non-stationary and noisy environment / phenomena (Rioul & Vellerli, 1991). Foufoula-Georgion and Kumar (1994) described the basic properties that make Wavelet analysis a powerful tool for geophysical applications.

In the last decade, wavelet transform (WT) has become a useful technique for analysing variations, periodicities, and trends in time series (Lu 2002; Coulibaly and Burn, 2004; Partal and Kucuk, 2006 a). A wavelet transformation is a strong mathematical signal processing tool like Fourier Transformation with the ability of analyzing both stationary as well as non stationary data, and to produce both time and frequency information with a higher resolution, which is not available from the traditional transformation (Fourier Transform and Short Time Fourier Transform). WT provides multi resolution analysis specifically at low scales (high frequency) it gives better time resolution and poor frequency resolution and at high scales (low frequency) it gives better frequency resolution and poor time resolution. In actual practice for all time series signals, such information is important. A non-stationary time series can be decomposed into certain number of stationary time series by WT. Then different single prediction methods are combined with wavelet transform to improve the prediction accuracy.

Wavelet theory (Mallat, 1989 a) was first developed in the end of 1980s of last century. In recent times, it has been applied in many fields, such as signal process, image compression, voice code, pattern recognition, hydrology, earthquake investigation and many other non-linear science fields. The researches and applications of wavelet analysis have already begun in hydrology and water resources. The document (Li et al., 1997) points out the potential applications of wavelet analysis to hydrology and water resources. Li et al. (1999) probed long time interval forecast of hydrological time series with combining neural network models with wavelet transform. Wang et al. (2000) had proposed a wavelet transform stochastic simulation model, which generated synthetic streamflow sequences that are statistically similar to observed streamflow sequences. The multitime scale characteristics of hydrological variables had been studied by Wang et al. (2002).

Wavelet analysis has been a hot research tool in time series analysis due to its multiresolution function (Zhou et al., 2008). Various studies demonstrated the effectiveness of this practice due to this ability of wavelet analysis (Wensheng and Jing, 2003; Nourani et al., 2008; Kisi, 2009 a; Goyal, 2013; Rathinasamy et al., 2013).

1.4 PROBLEM BACKGROUND

Identification of the multi-temporal scale characteristics of hydrologic series is very important for understanding complicated hydrologic processes (Labat, 2005; Neupauer et al. 2006), but conventional time series analysis methods cannot sufficiently meet this need. Wavelet coefficients under different scales reflect different characteristics of the series: positive coefficient values correspond to wet seasons and minus values correspond to dry season, and zero point corresponds to the turnover point from wet to dry season, or from dry to wet season; and absolute values of wavelet coefficients show the significance of characteristics: bigger absolute wavelet coefficient values reflect more statistical significance of the characteristics, and vice versa. Wavelet analysis has been widely and maturely applied in various geographic basins and regions worldwide for tackling the complex characteristics of hydrologic processes under multi-temporal scales. The multiresolution approach may be useful for the analysis of multi-scale features, detection of singularities, analysis of transient phenomena of non-stationary series and fractal processes (Salvatore et al., 2012).

Many new understandings and conclusions about the variability of hydrologic processes have been gained from the wavelet point of view. However, as a small shift in the input series would cause very different output wavelet coefficients (Chen and Xie, 2007), the stability of wavelet transform results should be further analyzed and understood.

Since last decade, many researchers observed that coupling wavelet transforms with ANN could improve hydrological forecast significantly (Partal & Cigizoglu, 2008; Zhou et al. 2008; Kisi, 2008, 2009; Adamowski & Karen, 2010). These studies found that coupled wavelet ANN models generally provided more accurate forecasts than other models. Empirical results show that networks trained with preprocessed data performed better than networks trained on un-decomposed, noisy raw data. In most of the hybrid models, WT is used as pre-processing technique. The wavelet-transformed data aid in improving the model performance by capturing helpful information on various resolution levels. Due the above mentioned advantages of WT, it has been found that the hybridization of wavelet transformation with other models like ANN, FL, ANFIS, linear models, etc., improved the results significantly compared to the single regular model (Deka and Prahlada, 2012).

Motivated by the effective pre-processing capability of wavelets and the predictive power of ANN models, it is proposed to evolve a unified framework for analyzing non-stationary time series. The proposed study will investigate the effectiveness of wavelet-based pre-processing for an ANN system for forecasting non-stationary hydrological time series. In addition, non-coupled ANN methods were used for comparison. Five forecasting horizons (2, 3, 4, 7, and 14 days ahead) for daily time series, two forecasting horizons (1, and 2 week ahead) for weekly data, and one forecasting horizon (1 month ahead) for monthly time series were considered. Also, the impact of various filter banks such as Haar and Daubechies of type db2, db3, db4, and db5 along with various decomposition levels on a single back propagation multilayered neural network for the prediction of daily, weekly and monthly discharge was investigated.

1.5 STUDY AREA

The Brahmaputra River in India forms a complex river system characterized by the most dynamic and unique water and sediment transport pattern. The river originates as Tsangpo, the source of which is at $31^{0}30'$ N and 82^{0}E in Tibet**,** in a great glacier mass in the Kailash range of the Himalayas to the east of the Mansarovar Lake (elevation 5300 m) and flows through China, India, and Bangladesh for a total distance of 2880 km before emptying into the Bay of Bengal through a joint channel with the Ganga. The basin lies between latitudes $24^{0}13'$ and $31^{0}30'$ North and longitudes 82^{0} and $96^{0}49'$ East**.** In Tibet, where it is called Tsangpo, the Brahmaputra flows eastward for 1100 km along the bottom of a longitudinal graben parallel to and about 160 km north of the Himalayas. At the extreme eastern end of its course in Tibet, the Tsangpo suddenly enters a deep narrow gorge at Pe (3500 m), which skirts around the Namcha Barwa peak (7755 m) and continued southward across the

Himalayan ranges. The gradient of the river in the gorge section ranges from about 4.3 m. to 16.8 m. per km. On entering India, the Tsangpo, now called Dihang, traverses 226 km of mountainous course before debouching onto the Assam plain near Pasighat (elevation 155 m.). At the exit of the gorge the slope of the river is only 0.27 m. per km. Near Kobo, 52 km. south of Pasighat, two rivers (Dibang and Lohit) meet the Dihang river, and the combined flow, called Brahmaputra, moves westward thorough Assam for 720 km. until near Dhubri, where it swerves to the south and enters Bangladesh. The Brahmaputra has a gradient of 0.09-0.17 m/km. near Dibrugarh at the head of the valley and is further reduced to about 0.1 m/km. near Guwahati (Pandu).

The Brahmaputra is the fourth largest river in the world in terms of average discharge at mouth, with a flow of 19,830 cumec (Goswami, 1985). The study area is located in the international river basin of Brahmaputra main stream within India. Pandu (u/s) and Pancharatna (d/s) stations are selected for the study. Ten year daily flow data for these two stations were collected from Water Resources Department, Assam, India. The catchment area upto Pandu station is 500,000 km^2 and upto Pancharatna it is 5,32,000 km^2. The annual rainfall in the Assam part is 2300 mm. The hydrologic regime of the river responds to the seasonal rhythm of the monsoons and to the freez-thaw cycle of the Himalayan snow. The rainy season (June to October) accounts for 82 % of the mean annual flow at the Pancharatna. The discharge is highly fluctuating in nature. Discharge per unit drainage area in the Brahmaputra Basin River is among the highest of major rivers of the world. Large variations of discharge within a short span of time are noticed during the flood season, with maximum difference of about 17000 cumecs in 24 hours (June 7-8, 1990) and 24000 cumecs in 48 hours (June 7-9, 1990) (Sharma J. N. 2005) being recorded in the rising limb. The locations of the two discharge gauging stations are shown in **Fig. 1.1** and the longitudinal profile of Brahmaputra River is shown in **Fig. 1.2.**

In this major river, due to the topographic restrictions, no major hydraulic structure exist as of now and no water diversion is done between the gauging sites.

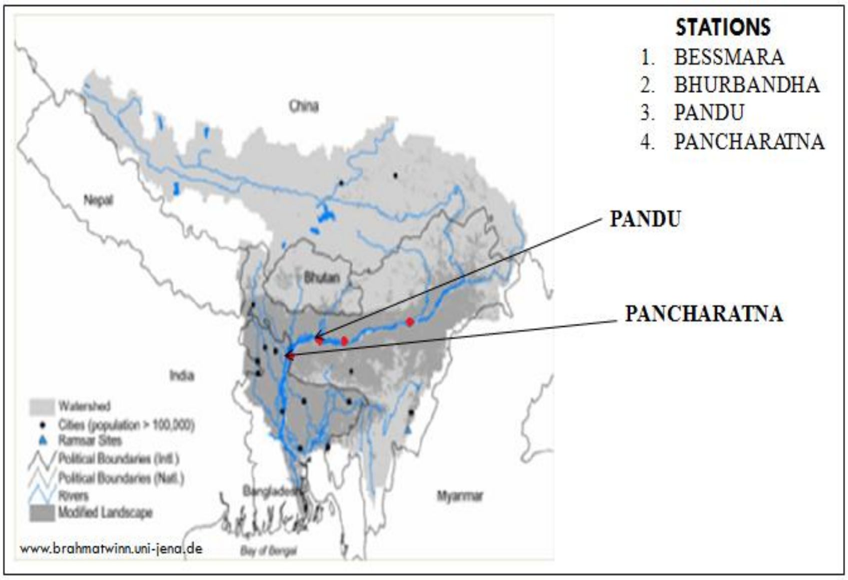


Fig.1.1 Location of the gauging sites

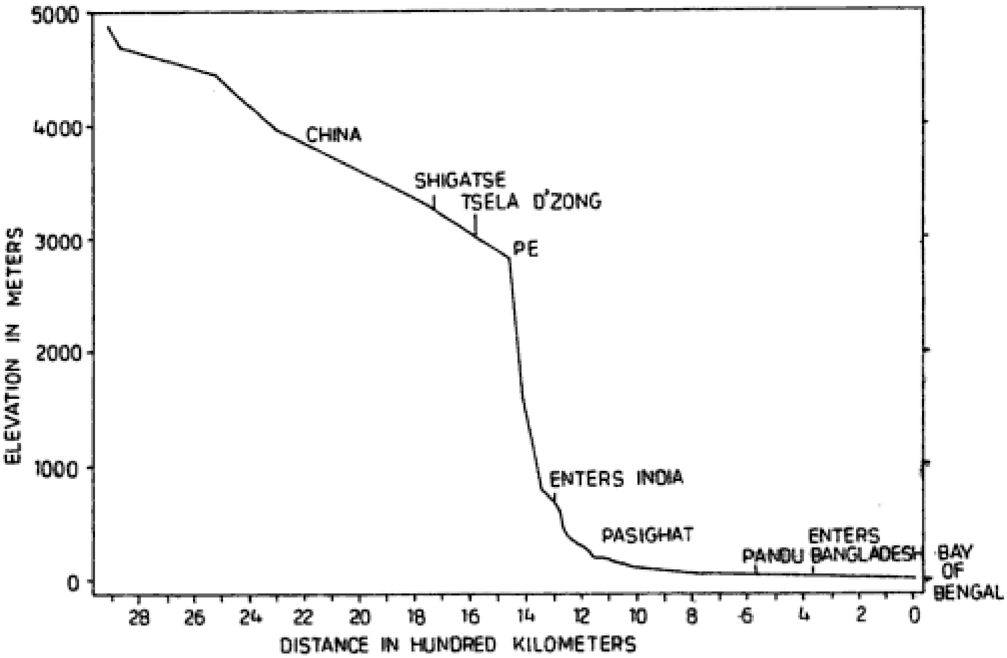


Fig. 1.2 Longitudinal profile of the Brahmaputra River (Goswami, 1985)

1.6 ORGANIZATION OF THE THESIS

This thesis comprises of five chapters as follows:

Chapter 1 Introduction: presents the relevant information pertaining to time series and further deals with overview of the conceptual basis for the research, and study area.

Chapter 2 Literature Review: deals with a brief discussion about the work carried out by previous researchers using ANN and hybrid WANN models for hydrologic time series forecasting and the objectives of the study.

Chapter 3 Methodology and Model Development: this chapter describes the basics of wavelet transform and artificial neural network. This also deals with the procedure for developing ANN and WANN models.

Chapter 4 Results and Discussion: this chapter describes the method of evaluation and goes on to present the analysis of the results obtained from the developed ANN and WANN models.

Chapter 5 Summary and Conclusions: represents summary of the research work carried out, contribution and conclusions. Further, the future scope is included towards the end.

CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

Streamflow, which is known an integrated process of atmospheric and topographic processes, is of prime importance to water resources planning. Recently, Artificial Intelligence (AI) techniques have shown great ability in modeling and forecasting non-linear hydrological time series incorporating machine learning in hydrology. Classical time series models such as ARMA (Autoregressive moving average) and ARIMA (Auto regressive integrated moving average) are basically linear models assuming that data are stationary, and have a limited ability to capture non-stationarities and non-linearity in data series. On the other hand, soft computing normally utilizes tolerance to uncertainties, imprecision, and partial truth associated with input information in order to come up with robust solution, handling non-linearities and non-stationarities effectively.

2.2 LITERATURE REVIEW

In the following sections the work carried out by various researchers in the field of Hydrology and Water Resources Engineering using artificial neural network (ANN), Wavelet –ANN (WANN) hybrid models, and wavelet techniques combined with other soft computing techniques like fuzzy logic (FL), genetic algorithm (GA), etc. are discussed in brief.

2.2.1 Using ANN Models

Christian et. al., (1998), studied ANNs for flow forecasting in two flood-prone UK catchments using real hydrometric data. Given relatively brief calibration data sets it was possible to construct robust models of 15-min flows with six hour lead times for the Rivers Amber and Mole. Comparisons were made between the performance of the ANN and those of conventional flood forecasting systems. The results obtained for validation forecasts were of comparable quality to those obtained from operational systems for the River Amber. The ability of the ANN to cope with missing data and to "learn" from the event currently being forecast in real time were observed.

Cigizoglu and Kisi (2005), low forecasting performance by artificial neural networks (ANNs) is generally considered to be dependent on the data length. In this study k-fold partitioning, a statistical method, was employed in the ANN training stage. The method was found useful in the case of using the conventional feed-forward back propagation algorithm. It was shown that with a data period much shorter than the whole training duration similar flow prediction performance could be obtained. Prediction performance and convergence velocity were compared between three different back propagation algorithms, Levenberg–Marquardt, conjugate gradient and gradient descent. The LM technique was found advantageous giving more satisfactory performance.

Wu J. S. et al., (2005) Used ANN for watershed-runoff and stream-flow forecasts conducted on a small urban watershed in Greensboro, North Carolina. Two ANN-hydrologic forecasting models for watershed runoff prediction model to predict storm water runoff at a gauged location near the watershed outlet and another stream flow forecasting model was formulated to forecast river flows at downstream. Results obtained from both model applications were very encouraging even with a relatively small number of storm events employed for training and testing.

Ozgur Kişi, (2007 a), studied using ANN's algorithms for short term daily streamflow forecasting. Four different ANN algorithms, namely, back propagation, conjugate gradient, cascade correlation, and Levenberg–Marquardt were applied to continuous streamflow data of the North Platte River in the United States. The models were verified with untrained data. The results from the different algorithms were compared with each other. The correlation analysis was used in the study and found to be useful for determining appropriate input vectors to the ANNs.

Ozgur Kişi., (2007 b), investigated the abilities of range-dependent neural networks (RDNN) to improve the accuracy of streamflow-suspended sediment rating curve in daily suspended sediment estimation. A comparison was made between the estimates provided by the RDNN and those of the following models: Artificial neural networks (ANN), linear regression (LR), range dependent linear regression (RDLR), sediment rating curve (SRC) and range-dependent sediment rating curve (RDSRC). The daily streamflow and suspended sediment data belonging to two stations-Calleguas Station and Santa Clara Station operated by the US Geological Survey were used as case studies. Based on comparison of the results, it was found that the RDNN model gave better estimates than the other techniques. RDLR technique was also found to perform better than the single ANN model.

Surinder et al., (2008*),* studied an ANN based modeling technique to study the influence of different combinations of meteorological parameters on evaporation from a reservoir. Several input combination were tried so as to find out the importance of different input parameters in predicting the evaporation. The prediction accuracy of Artificial Neural Network had also been compared with the accuracy of linear regression for predicting evaporation. The comparison demonstrated superior performance of ANN over linear regression approach. The highest correlation coefficient (0.960) along with lowest root mean square error (0.865) was obtained with the input combination of air temperature, wind speed, sunshine hours and mean relative humidity. The findings of this study suggest the usefulness of ANN technique in predicting the evaporation losses from reservoirs.

Solaimani K. (2009), utilized ANN for modeling the rainfall runoff relationship in a catchment area located in a semiarid region of Iran by adopting feed forward back propagation for the rainfall forecasting with various algorithms with performance of multi-layer perceptions. The monthly stream flow of Jarahi Watershed was analyzed in order to calibrate of the given models. The monthly hydrometric and climatic were ranged from 1969 to 2000. The results extracted from the comparative study indicated that the ANN was more appropriate and efficient to predict the river runoff than classical regression model.

Zadeh et al. (2010), studied (ANN) models for predicting daily flows from Khosrow Shirin watershed located in the northwest part of Fars province in Iran. A Multi-Layer Perceptron (MLP) neural network was developed using five input vector using 5-year data record adopting Levenberg–Marquardt (LM) algorithm. It was found that antecedent precipitation and discharge with 1 day time lag as an input vector best predicted daily flows. Also, comparison of MLPs showed that an increase in input data was not always useful. The predicted outflow showed that the tangent sigmoid activation function performed better than the logistic sigmoid activation function.

Besaw et al., (2010) studied two artificial neural networks (ANNs) to forecast streamflow in ungauged basins. The model inputs include time-lagged records of precipitation and temperature. In addition, recurrent feedback loops allow the ANN streamflow estimates to be used as model inputs. Streamflows from sub-basins in Northern Vermontare were used to train and test the methods. To predict streamflow in an ungauged basin, the recurrent ANNs were trained on climate-flow data from one basin and used to forecast streamflow in a nearby basin with different (more representative) climate inputs. One of the key results of this work was that these recurrent flow predictions were being driven by time-lagged locally-measured climate data. A scaling ratio, based on a relationship between bank full discharge and basin drainage area, accounted for the change in drainage area from one basin to another. Hourly streamflow predictions were superior to those using daily data for the small streams tested due to loss of critical lag times through up scaling. The ANNs selected in this work always converged, avoiding stochastic training algorithms, and are applicable in small ungauged basins.

Mehmet et al. (2009), carried out study on the issue of flow forecast based on the soil and water assessment tool (SWAT) and artificial neural network (ANN) models. In this study, the ANNs were applied to the daily flow of the Pracana basin in Portugal. A comparison of ANN models and a process- based model SWAT was conducted based on their prediction accuracy. The ANN model was found to be more successful than the SWAT in relation to better forecast of peak flow. The SWAT model results revealed a better value of mean squared error. The study revealed that ANNs can be powerful tools in daily flow forecasts.

Asadi S. et al. (2013), proposed a hybrid intelligent model for runoff prediction of Aghchai watershed. The model was a combination of data pre-processing methods, genetic algorithm and Levenberg-Marquardt (LM) algorithm for learning feed forward neural networks. The authors used data pre-processsing methods such as data transformation, input variables selection and data clustering for improving accuracy. The results showed that the prediction of runoff was more accurate than ANN, and adaptive neuro fuzzy inference system (ANFIS) models.

2.2.2 Using Hybrid Models of WANN

The wavelet-ANN is another reliable hybrid model used in time series forecasting problems. Recently, wavelet transform analysis has become a popular analysis tool due to its ability to analyse simultaneously both spectral and temporal information within the signal. Some of the recent works carried out in Hydrology are discussed below.

Addison et al. (2001), used wavelet transform analysis to a variety of open channel wake flows. Feature location was undertaken using a continuous wavelet transform, and both turbulent statistical analysis and thresholding of the turbulent signal components were undertaken using a discrete wavelet transform. It was found that the continuous wavelet transform was the preferred method for feature detection within fluid velocity time signals.

Wensheng Wang et al., (2003), carried out a multi-time scale prediction of ground water level at Beijing and daily discharge of Yangtze River Basin at China using Hybrid Model of Wavelet-Neural Network. Through a Trous algorithm and three-layer neural network forecasting results were carried out. 12 years of shallow monthly ground water level data were used, 9 years for calibration and 3 years for validation. Daily discharge data of 8 years were used for training and 2 years for testing. The comparisons revealed that the model increased the forecasted accuracy and prolonged the length time of prediction. The proposed WANN model focused on improving the precision and prolonging the forecasting time period.

Kim and Valdes, (2003) developed nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks in the Conchos river basin in Maxico. The results indicated that the conjunction model using dyadic wavelet transform significantly improved the ability of neural network in forecasting.

Cannas B. et al., (2006*),* studied the river-flow forecasting one month ahead with Neural Networks and Wavelet Analysis using monthly runoff data for the Tirso Basin, Italy. The data set was split into three parts- first 40 years was used for training, next 9 years for cross validation and last 20 years for testing. The reconstruction of the data was done by traditional feed forward, MLP networks. For the non-stationary and seasonal irregularity of runoff time series, the best results were obtained using data clustering and discrete wavelet transform combination. Tests showed that neural networks trained with pre-processed data showed better performance.

Zhou et al. (2008) developed monthly discharge predictor-corrector model based on wavelet decomposition using 52 years records of monthly discharge at Yichang station of Yangtze River**.** The decomposed times series data were used as input to ARMA model for prediction which improved the prediction accuracy**.**

Rao et al., (2009), carried out modelling using hydrological time series data adopting Wavelet-Neural Network for four west flowing rivers in India namely Kollur,(22 years data from 1981-2002), Seethanadi (26 years data from 1973-1998), Varahi(26 years 1978-2003) and Gowrihole (25 years data from 1979-2003). The models of WANN having different antecedent values of the time series showed minimum RMSE (0.88 to 0.54 m), high correlation coefficient (0.93 to 0.96) and highest efficiency (87.21 to 92.86 %) during validation period. The results of daily streamflow and monthly groundwater level series modelling indicated that the performances of WANN Models were more effective than ANN Models. He recommended that the models developed for streamflow and groundwater levels would be useful for water resources planning in the Western Ghats and for ground water management in coastal aquifers.

Nourani et al., (2009 a), studied the rainfall-runoff modelling using Wavelet - ANN approach for predictions of runoff discharge one day ahead of the Ligvanchai watershed at Tabriz, Iran. The daily rainfall and runoff time series for 21 years were used. The time series were decomposed upto four levels by using Haar, Daubechies (db2), Symlet (sym3) and Coiflet (coif1). The Study showed that both short and long term runoff discharges could be predicted satisfactorilly. The model results showed the high merit of Haar wavelet in comparison with the others. They also recommended that Wavelet Transform could be used for trend analysis in watersheds.

Kisi, O. (2009 a), developed neurowavelet (NW) model by combining two methods discrete wavelet transform (DWT) and artificial neural network (ANN), for 1 day ahead intermittent streamflow forecasting and results were compared with those of the single ANN model. Intermittent streamflow data from two stations in the Thrace Region, the European part of Turkey, in the northwest part of the country were used in the study. In NW model, the original time series were decomposed into a five subtime series components by Mallat DWT algorithm. The correlation coefficients between each subtime series and original intermittent streamflow time series were found. These correlation values provide information for the determination of effective wavelet components on streamflow. The new subtime series having high correlation coefficient were used as input to the ANN model. The NW model was found to be much better than the ANN in high flow estimation. The test results showed that the DWT could significantly increase the accuracy of the ANN model in modelling intermittent streamflows.

Rajaee T. et al. (2010), investigated the prediction of daily suspended sediment load one day ahead with wavelet and neuro-fuzzy combination model using time series data of discharge and suspended sediment load as input in a gauging station from the Pecos River in USA. Results showed that the wavelet analysis and neuro-fuzzy (WNF) model performed better predictions rather than neuro-fuzzy and sediment rating curve. The cumulative suspended sediment load estimated by this technique was closer to the actual data. The WNF model considered periodic and stochastic characteristics of suspended sediment phenomenon and may provide suitable constrictions not clearly seen in the suspended sediment rating curve. The model also could be employed to stimulate hysteresis phenomenon, while the sediment rating curve method is incapable in this event.

Adamowski and Karen, (2010), investigated a method based on coupling discrete wavelet transform (WA) and ANN for flow forecasting applications in non-perennial rivers in semi-arid watersheds at lead times of 1 and 3 days for two different rivers in Cyprus. The discrete trous wavelet transform was used to decompose flow time series data into 8 levels wavelet coefficients which were used as inputs to Levenberg Marquardt artificial neural network models to forecast flow. WA-ANN model provided more accurate results than regular ANN.

Shiri and Kisi, (2010), studied short-term and long term streamflow forecasting using a wavelet and neuro-fuzzy conjunction model to investigate the daily, monthly and yearly streamflows at Derecikviran station on Filyos River in the Western Black Sea region of Turkey using 31 years of streamflow data. The results obtained showed that the neuro-fuzzy (NF) and wavelet-neuro-fuzzy (WNF) models increased the accuracy of the single NF models especially in forecasting yearly streamflow. Also the single NF and WNF models were compared with each other by adding periodicity components into the inputs. The comparison of the results indicated that adding periodicity component generally increased the model's accuracy.

Kisi, O. (2010) developed neuro-wavelet models for daily suspended sediment estimation for two stations on Tongue River in Montana using daily streamflow and suspended sediment data. The comparison of results revealed that the developed model could increase the estimation accuracy.

Adamowski and Chan . (2011), developed hybrid wavelet transforms-neural network for monthly groundwater level forecasting using monthly total precipitation, monthly average temperature and average ground water levels as input variables. The original data was decomposed into a series of details using a modified version of a trous DWT. The WANN models were found to provide more accurate monthly average ground water level forecasts compared to the ANN and ARIMA models.

Nourani et al., (2011*)***,** studied two hybrid artificial intelligence (AI) models for two watersheds located in Azerbaijan, Iran, for modeling rainfall–runoff process. Two hybrid AI-based models which were reliable in capturing the periodicity features of the process were introduced for modeling. In the first model, the SARIMAX (Seasonal Auto Regressive Integrated Moving Average with exogenous input)-ANN model, an ANN was used to find the non-linear relationship among the residuals of the fitted linear SARIMAX model. In the second model, the wavelet-ANFIS model, wavelet transform was linked to the ANFIS concept and the main time series of two variables (rainfall and runoff) were decomposed into some multi-frequency time series by wavelet transform. Afterwards, these time series were imposed as input data to the ANFIS to predict the runoff discharge one time step ahead. The obtained results showed that, although the proposed models can predict both short and long terms runoff discharges by considering seasonality effects, the second model was relatively more appropriate because it used the multi-scale time series of rainfall and runoff data in the ANFIS input layer.

Kisi and Shiri (2011) developed precipitation forecasting model using wavelet-genetic programming and wavelet-neuro-fuzzy conjunction. They found that hybrid wavelet-genetic programming model was of better performance than hybrid wavelet-neuro-fuzzy model.

Rajaee T. et al. (2011), developed artificial neural network (ANN), wavelet analysis and ANN combination (WANN), multilinear regression (MLR), and sediment rating curve (SRC) models for daily suspended sediment load (*S*) modeling in the Iowa gauging station in the US. In the WANN model, discrete wavelet transform was linked to the ANN method. For this purpose, the observed time series of river discharge (*Q*) and *S* were decomposed into 5 levels by discrete wavelet transform (DWT) which were imposed as input to ANN to predict one day ahead *S*. A complex Morlet wavelet technique was applied to analyze wavelet construction of daily *Q* and *S*. The number of nodes in the input in WANN model was determined by $(i + 1) \times 2$, because this model uses two variables (*Q* and *S*) and each time series is decomposed into i, $i = (1,2,\ldots,5)$ detailed time series and approximation time series.

This study was aimed at examining the effects of the employed mother wavelet type on the proposed WANN model efficiency. Seven different mother wavelets were used [viz. Daubechies-2 (db2) (the most popular wavelet), the Haar wavelet (a simple wavelet), and some irregular wavelets such as Bior1.1, Rboi1.1, Coif1, Sym1, and Mayer wavelets].

It was found that, increasing the decomposition level, in levels over Level 1, decreases the model's performance, because high decomposition levels lead to a large number of parameters with complex nonlinear relationships in the ANN technique. The WANN model was more accurate in predicting the *S* and its performance was better than the ANN, MLR, and SRC models.

Wang et al. (2011), utilized wavelet transform method for synthetic generation of daily streamflow in Jinsha river of China. Daily streamflow sequences with different frequency components were decomposed into the series of wavelet coefficients at various resolution levels using wavelet decomposition algorithm. Based on these sampled subseries, a large number of synthetic daily streamflow sequences were obtained using wavelet reconstruction algorithm. They concluded that this newly developed method was able to assess the of probability distributions type and of dependence structure.

Abghari et al. (2012), developed wavelet neural network hybrid model for prediction of daily pan evaporation using Mexian hat and polyWOGI mother wavelets. Results showed that Mexican hat wavelet neural network in the best topology presents 96.04 % accuracy, while polyWOGI wavelet neural network presents 91.03 % accuracy in testing period. The MLP model with standard sigmoid function resulted in 87.63 % accuracy in testing period. Comparison showed Mexican hat neural network had better accuracy.

Nayak et al. (2013), developed wavelet neural network (WNN) hybrid model for Malprabha river (India) flow forecasting using rainfall, runoff and evaporation as inputs to ANN and WNN models. They used Daubechies wavelet of order 5 (db5) for one time step ahead forecasting. Results of this study had been compared by developing standard ANN and NAM models. Results indicated that the WANN model performed better than ANN and NAM models.

Krisnna B. (2013). Main purpose of the study was to examine the capability of two pre-processing techniques such as wavelets and moving average (MA) methods in combination with feed forward neural networks and multiple linear regression (MLR) in the prediction of daily inflow values of Malprabha river, India. Two NN structures namely back propagation (BP) and radial basis (RB) algorithms were used. Daily data of rainfall, reservoir inflow and discharge for 11 years were used in the study. The author found that optimal input combination based on the method suggested by Sudheer et al. (2002). The time series data were decomposed into 3 levels using mother wavelets db5, db1, db2, db6, Bior1.1, Coif 1, Meyer, rbio1.1 and sym3. Results indicated that the hybrid WNN model performed better compared to ANN and MLR models.

Nourani et al. (2013), in this research, a two level self organizing map (SOM) clustering technique was used to identify spatially homogeneous clusters of precipitation satellite data, and to choose the most operative and effective data for a feed forward neural network (FFNN) to model rainfall-runoff process on daily and multi-step (2, 3 and 4 day) ahead time scale. The WT was used to extract dynamic and multi-scale features of the non-stationary runoff time series and to remove noise. The authors used db4 and Haar wavelets. The performance of the coupled SOM-FFNN model was compared to the newly proposed SOM-WT-FFNN model. Also these two models were compared with auto regressive integrated moving average with exogeneous input (ARIMAX) model. It was found that the application of WT to the runoff data increased the performance of the FFNN rainfall-runoff models in predicting runoff peak values by removing noise. Also, it was found that db4 wavelet was superior to Haar wavelet for runoff forecasting.

Karthikeyan and Nagesh Kumar (2013), in this paper the predictability of wavelet based and Empirical Mode Decomposition (EMD) based time series modeling techniques were studied under various case studies of monthly total stream flow (4 non-stationary sites) and monthly total rainfall (two non stationary sites) locations. The time series data was decomposed using these two techniques. The predictability was checked for six and twelve months ahead forecasts. It was observed that the wavelet based method has better prediction capability over EMD.

Other than the works mentioned above, some other researchers had developed rain-fall runoff models using wavelet – artificial neural network (WANN) (Adamowski J., 2008; Adamowski et al., 2009; Chou, 2004; Kisi, 2011; Kisi et al., 2013; Venkata Ramana et al., 2013). Applications of WANN models have been found in other fields like wave height forecasting, urban water demand forecasting, drought, evapotranspiration, etc in the literature (Campisi et al., 2012; Deka et al., 2010; Prahlada and Deka, 2011; Partal, 2009; Shirmohammadi et al., 2013). Recently wavelet transformations had been combined with other soft computing techniques like regression, fuzzy logic, genetic algorithm, support vector machine (Kisi, 2011; Kisi and Cimen, 2012; Partal and Kisi, 2007; Wang et al. 2011). Maheswaran and Khosa (2012) studied comparative study of different wavelets for hydrologic forecasting, Sang and Wang (2008) studied wavelets selection method in hydrologic series wavelet analysis, Sang Y. F. (2012) has given a practical guide to discrete wavelet decomposition of hydrologic time series.

2.3 SUMMARY OF LITERATURE REVIEW AND RESEARCH OBJECTIVES

Most of the previous studies were carried out using selective type of mother wavelet. Also, the effect of decomposition level on model forecasting ability for stream flow forecasting was not well documented so far. Again the forecasting accuracy for multi-step lead-time using short term and long term time series data was not addressed comprehensively so far. The potential of Daubechies wavelet of various orders in analysing hydrologic time series behavior has been explored thinly so far, that require further threadbare study.

Keeping all this in mind, it is proposed to develop a hybrid model combining wavelet and artificial neural network with the following objectives:

1. To investigate the potential and forecasting ability of the hybrid model by combining Wavelet - Artificial Neural Network (WANN) using time series data of various temporal scales.
2. Development of various WANN hybrid models for multistep lead times and their performance evaluated and results compared with single ANN models.
3. To assess the influence of different decomposition levels for various short term as well as long term flow forecasting on the model performance.
4. To investigate the proper selection of mother wavelets within the domain of Daubechies wavelet of various orders for better forecasting accuracy.

CHAPTER 3

METHODOLOGY AND MODEL DEVELOPMENT

This chapter is divided into two parts A) Methodology and, B) Model development. In the first part, in this research, hybrid models combing Wavelet Transform (WT) and Artificial Neural Network (ANN) are developed, the basics of WT and ANN are discussed, while in the second part the procedure for model development is discussed.

3.1 METHODOLOGY

Wavelet theory has been applied in many fields, such as signal process, image compression, voice code, pattern recognition, hydrology, earthquake investigation and many other non-linear science fields. Wavelet theory is discussed thoroughly in Labat et al. (2000) and Mallat (1998).

Signals whose frequency content does not change with time are called stationary signals. In other words, the frequency content of stationary signals does not change in time. In stationary signals it is not necessary to know at what times frequency components exist, since all frequency components exist at all times.

Mathematical transformations (viz. Fourier transform (FT), Short Time Fourier transform (STFT), Wavelet Transform (WT), etc. are applied to time domain signals (raw signals) to obtain further information from that signal that is not readily available in the raw signals. The above mentioned mathematical transformation techniques are briefly described in the following sections.

3.1.1 Fourier Transform (FT)

If the FT of a signal in time domain is taken, the frequency-amplitude representation of that signal is obtained. That is we have a plot with one axis being the frequency and the other being the amplitude. This plot tells us how much of each frequency exists in the raw signal. But it doesn’t tell about what spectral component exists at any given time instant i.e. the time information is lost. So FT is not suitable for non-stationary data. The FT is defined by the following two equations:

$$F(\omega) = \int_{-\infty}^{\infty} x(t).e^{-2j\pi\omega t} dt \tag{3.1}$$

$$x(t) = \int_{-\infty}^{\infty} F(\omega).e^{2j\pi\omega t} d\omega \tag{3.2}$$

In the above equation ω stands for frequency, *t* stands for time and *x(t)* denotes time domain signal. Equation (3.1) is FT of *x(t)* and equation (3.2) is inverse FT of *F(* ω *).* In Eq. (3.1), the signal *x(t),* is multiplied with an exponential term, at some certain frequency 'ω', and then integrated over all the times. This integral is calculated for every value of 'ω'. If the value of this integration is large, then this means that the signal has a major component of 'ω' in it.

3.1.2 Short Time Fourier Transform (STFT)

The STFT is an improvement on the FT because it provides a measure of time and frequency resolutions. The difference between STFT and FT is that in STFT, the signal is divided into small enough segments, where these segments (portions) of the signal can be assumed to be stationary. For this purpose, a window function '***w***' is chosen. The width of this window must be equal to the segment of the signal where its stationarity is valid. This window is first located at the very beginning of the signal. The window function and signal are then multiplied. This product is assumed to be another signal, whose FT is to be taken. In other words, FT of this product is taken, just like taking FT of any signal. The next step is shifting this window to new location, multiplying with the signal, and taking FT of the product. This procedure is followed until the end of the signal is reached.

STFT is defined as

$$STFT(t,\omega) = \int_{t} [x(t).w^{*}(t)].e^{-2j\pi\omega t}.dt \tag{3.3}$$

In the above equation *x(t)* denotes raw signal, *w(t)* denotes window function, and * is complex conjugate. Wide window gives good frequency resolution, but poor time resolution. Narrow window gives good time resolution, but poor frequency resolution.

The use of a fixed window size at all times and for all frequencies is a limitation of this method.

In short, FT of a signal in time domain gives information about how much of each frequency exists in the raw signal without giving the information about time (Misiti et al.2010). So FT is not suitable for non-stationary data. On the other hand, STFT provides a measure of time and frequency resolutions, but the use of a fixed window size at all times and for all frequencies is a limitation of this method.

3.1.3. Wavelet Transformation

The wavelet representation addresses the above limitation, by adaptively partitioning the time-frequency plane, using a range of window sizes. At high frequencies, the wavelet transform gives up some frequency resolution compared to the Fourier transform. **Figure 3.1** shows representation of the effect of using FT and WT. WT provides multi resolution analysis i.e. at low scales (high frequency) it gives better time resolution (represented by compact width of time window as shown in **Fig.3.1**) and poor frequency resolution (represented by wider width of scale window as shown in **Fig.3.1**) and at high scales (low frequency) it gives better frequency resolution and poor time resolution and in actual practice for all the time series signals such information is important. The lower scales (i.e. compressed wavelet) trace the abrupt change or high frequency of a signal and the higher scales (i.e. stretched wavelet) trace the slowly progressing occurrences or low-frequency component of the signal. The wavelet transform breaks the signal into its wavelets (small wave) which are scaled and shifted versions of the original wavelet so called mother wavelet.

The wavelet transformation is divided into two types:

1) Continuous wavelet transform (CWT)
2) Discrete wavelet transform (DWT)

3.1.3.1 Continuous wavelet transform (CWT)

The basic objective of the CWT is to achieve a complete time-scale representation of localized and transient phenomenon occurring at different time scales (Labat, 2008). The Continuous Wavelet Transform (CWT) of a signal *x(t)* is given by the Eq. 3.4.

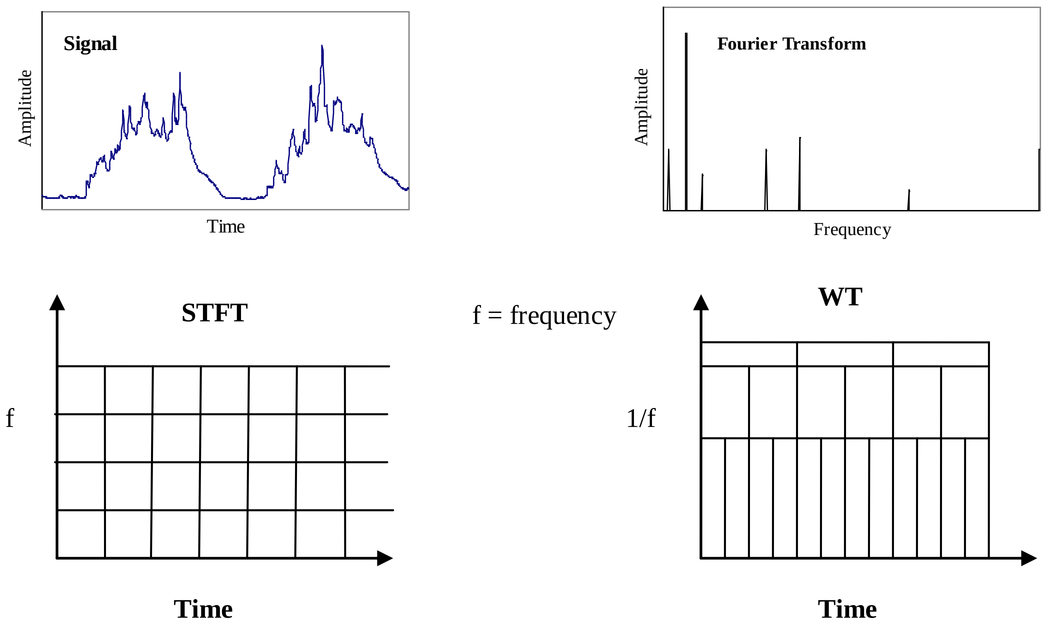


Fig 3.1. Fourier Transform and Wavelet Transformation

$$CWT(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t).\psi^{*}\left(\frac{t-b}{a}\right).dt \qquad (3.4)$$

In the above equation, the transformed signal is a function of two variables, *a* and *b*, the scale and translation factors, respectively, of the function ψ *(t)*. * corresponds to complex conjugate (Mallat, 1998). ψ *(t)* is the transforming function, and is called the mother wavelet, which is defined mathematically as

$$\int_{-\infty}^{\infty} \psi(t)dt = 0 \qquad (3.5)$$

The term translation is related to the location of the window, as the window is shifted through the signal. This term, obviously, corresponds to time information in the transform domain. The scale parameter is defined as 1/frequency. Low frequencies (high scales) correspond to a global information of a signal (that usually spans the entire signals), whereas high frequencies (low scales) correspond to a detailed information of a hidden pattern in the signal (that usually lasts a relatively short time).

The CWT is computed by changing the scale of the analysis window, shifting the window in time, multiplying by the signal, and integrating over all times.

The original signal is reconstructed using the inverse wavelet transform as

$$x(t) = \frac{1}{C_\psi} \int_{-\infty}^{\infty}\int_{0}^{\infty} \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) CWT(a,b) \frac{da.db}{a^2} \tag{3.6}$$

where C_ψ is admissibility constant.

The generation of wavelet coefficients for a time series involves five steps (Misiti et al.2010):

i) Given a signal *x(t)* and a wavelet function *Ψ(t)*, compare the wavelet to a section at the start of the signal. (**Fig. 3.2 a**).

ii) Compute the coefficient, (say *C1,1*; scale = 1, time = 1), which is an indication of the correlation of the wavelet function with the selected section of the signal.

iii) Shift the wavelet to the right (and find coefficient *C1,2*; scale = 1, time = 2) and repeat steps (i) and (ii) until the entire signal is covered. (**Fig. 3.2 b**).

iv) Dilate (scale) the wavelet (and find coefficient *C2,1*; scale = 2, time = 1) and repeat steps (i) through (iii). (**Fig. 3.2 c**).

v) Repeat steps (i) through (iv) for all scales to obtain coefficients at all scales and at different sections of the original signal.

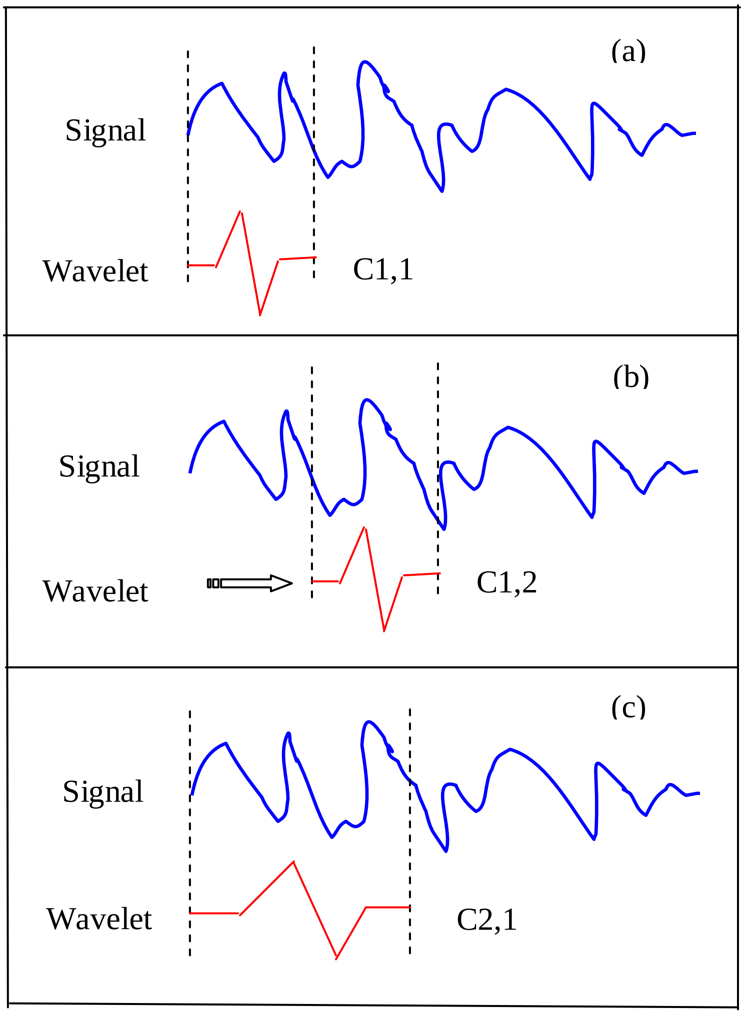


Fig. 3.2. Generating wavelet coefficients from a time series

3.1.3.2 Discrete wavelet transform (DWT)

Usually all hydrological times series data are observed at discrete time intervals, rather than continuous time. A discretization of Eq. 3.4 based on the trapezoidal rule is the simplest discretization of the continuous wavelet transform. Calculating the CWT coefficients at every possible scale is a fair amount of work, and it generates a lot of data. CWT produces N^2 coefficients from a data set of length N. Hence redundant information is locked up within the coefficients, which may or may not be a desirable property (Rajaee T. et al., 2011, Nourani et al., 2009 b). If one chooses scales and positions based on the powers of two (dyadic scales and positions) then the analysis will be much more efficient as well as accurate, which will provide

N transform coefficients. This transform is called discrete wavelet, and has the form as

$$\psi_{m,n}(t) = \frac{1}{\sqrt{a_o^m}}\psi\left(\frac{t - nb_o a_o^m}{a_o^m}\right) \tag{3.7}$$

where m and n are integers that control the wavelet dilation and translation, respectively; b_o is the location parameter and must be greater than zero; a_o is a specified fixed dilation step greater than 1. From this equation, it can be seen that the translation step $nb_oa_o^m$ depends upon the dilation, a_o^m. The most common and simplest choice for parameters a_o and b_o are 2 and 1 (time steps), respectively. This power of two logarithmic scaling of the translations and dilations is known as the dyadic grid arrangement (Mallat, 1989 b) and is defined as

$$\psi_{m,n}(t) = 2^{-m/2}\psi(2^{-m}t - n) \tag{3.8}$$

For discrete time series, x_t, where x_t occurs at discrete time t, the discrete wavelet transform becomes

$$W_{m,n} = 2^{-m/2}\sum_{i=0}^{N-1}\psi(2^{-m}t - n)x_t \tag{3.9}$$

where $W_{m,n}$ = wavelet coefficient for the discrete wavelet of scale $a = 2^m$ and location $b = 2^m n$. Eq. (3.7) considers a finite time series, x_t, $t = 0, 1, 2,\ldots, N - 1$, and N is an integer power of 2: $N = 2^M$; n is time translation parameter. This gives the range of n and m as, respectively, $0 < n < 2^{M-m} - 1$ and $1 < m < M$. At the largest wavelet scale (i.e. 2^m where $m = M$) only one wavelet is required to cover the time interval, and only one coefficient is produced. At the next scale (2^{m-1}), two wavelets cover the time interval, hence two coefficients are produced, and so on down to $m = 1$. At $m = 1$, the a scale is 2^1, i.e. 2^{M-1} or $N/2$ coefficients are required to describe the signal at this scale. The total number of wavelet coefficients for a discrete time series of length $N = 2^M$ is then $1 + 2 + 4 + 8 + \ldots + 2^{M-1}$ = N-1.

In addition to this, a signal smoothed component, $\overline{W}$, is added, which is the signal mean. Thus, a time series of length N is broken into N components, i.e. with zero redundancy. The inverse discrete transform is given by (Mallat, 1998):

$$x_t = \overline{W} + \sum_{m=1}^{M}\sum_{n=0}^{2^{M-m}-1} W_{m,n}2^{-m/2}\psi(2^{-m}t - n) \tag{3.10}$$

or in a simple format as

$$x_t = \overline{W}(t) + \sum_{m=1}^{M} W_m(t) \tag{3.11}$$

where $\overline{W}(t)$ is the approximation sub-signal at level M and $W_m(t)$ are the detail sub-signals at levels m = 1, 2,…., M. The detail wavelet coefficients, $W_m(t)$ can capture small features of interpretational value in the data. The residual term $\overline{W}(t)$ represents background information of the data.

DWT operates two sets of functions viewed as high-pass (wavelet function) and low-pass (scaling function) filters. The original time series are passed through high-pass and low-pass filters (as shown in **Fig. 3.3**) and down sampled by two (i.e throwing away every second data point) (Deka and Prahalada, 2012). After passing the signal through high pass and low pass filters, detailed (D_1, D_2,…., D_n, which are high frequency components of the original signal) and approximation coefficients (A_1, A_2,….A_n, which are low frequency components of the origininal signal), respectively, are obtained. At any n^{th} decomposition level there will be one series of approximation coefficients at n^{th} level (i.e. A_n) and n series of detailed coefficients (i.e. D_1, D_2,…., D_n), hence there will be total n+1 coefficients and the sum of $A_n + D_1 + D_2 + \ldots\ldots + D_n$ is equal to the original signal x(t).

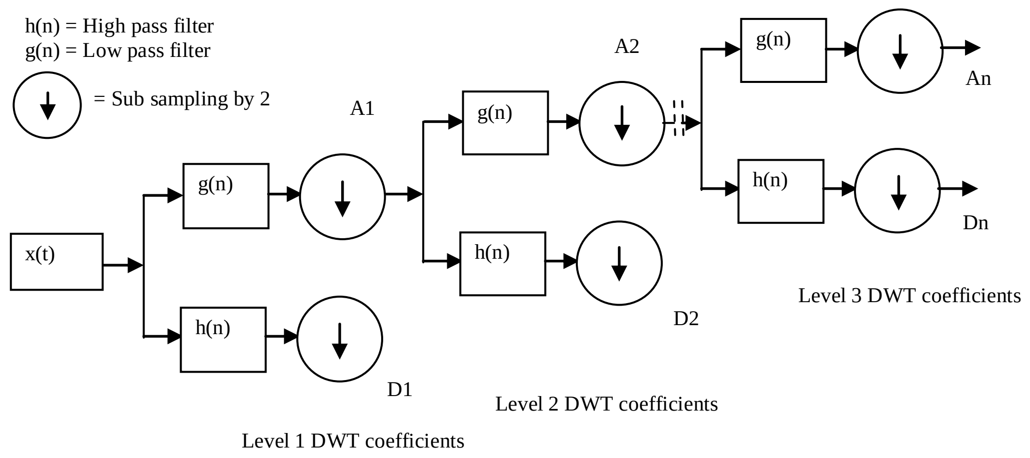


Fig.3.3. Wavelet decomposition tree

After wavelet decomposition it is required to reconstruct the singnal. The process of asembling back into original signal without loss of information is called reconstruction or synthesis. In wavelet reconstruction the decomposed approximation and detail doefficients are upsampled by two and then they are passed through low pass and high pass filters, respectively, as shown in Fig. 3.4 to get the original signal. The wavelet reconstruction process consists of upsampling and filtering. Upsampling is the process of lengthing a signal component by inserting zeros between samples.

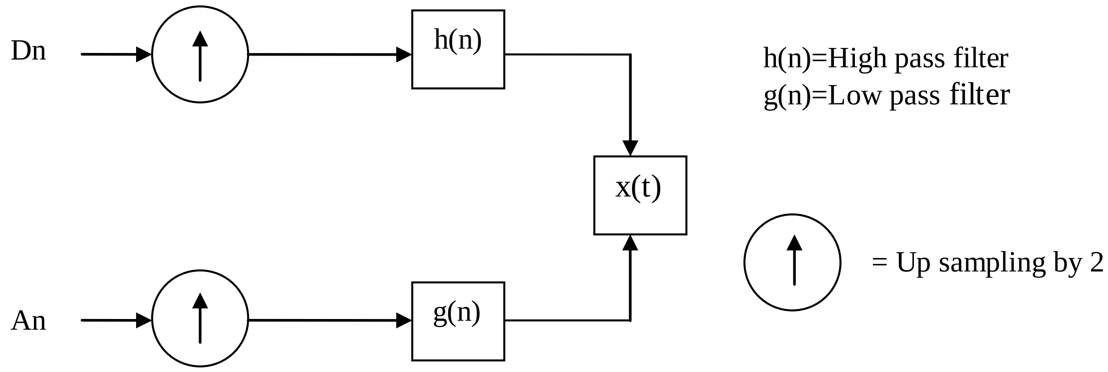


Fig. 3.4 Wavelet reconstruction tree

3.1.4 Artificial Neural Network (ANN)

Definition of ANN According to Haykin (1998): A neural network is a massively parallel distributed processor that has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects:

–Knowledge is acquired by the network through a learning process.

–Inter neuron connection strengths known as synaptic weights are used to store the knowledge.

ANN has gained popularity among Hydrologist in recent decades due to its large array of applications in the field of Engineering and research. The first neuron was produced in 1943 by the neurophysiologist Warren McCulloch and the logician Walter Pits. Thereafter, till 1969 Minsky and Papert wrote a book in which they generalized the limitations of Artificial Neural Networks. The era of renaissance started with John Hopfield in 1984 introducing recurrent neural network architecture.

Neural network is an inter connected group of artificial neurons, that can be used as computational model for information processing based on connectionist approach to computation. These are non–linear statistical data modeling tools, which can be used as model to develop a good relationship between input and output.

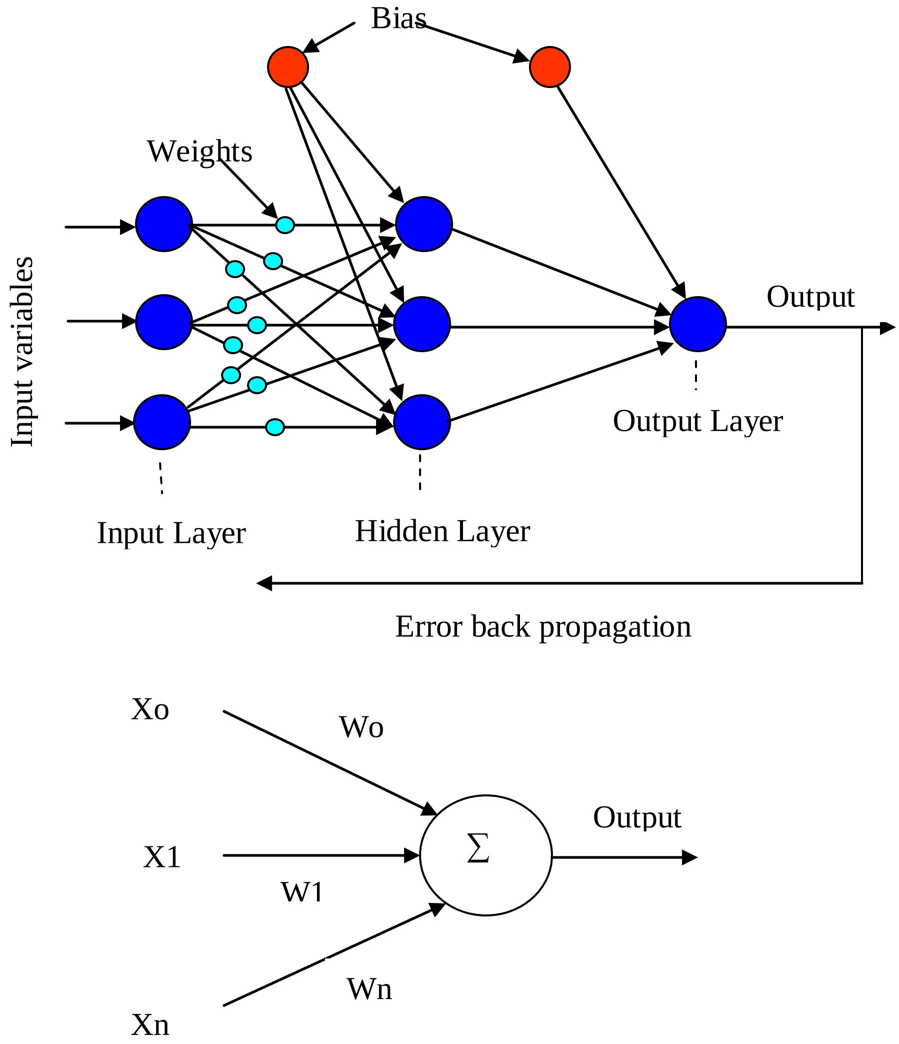
Mathematically, an ANN can be treated as universal approximators having an ability to learn from examples without the need of explicit physics.

In most of hydrologic time series models three layer-feedforward (**Fig. 3.4**) type of artificial neural network is used (Jain and Chalisgaonkar, 2000; Raghuwanshi et al., 2006; Tayfur, 2006). In the present study, feedforward ANN with Lavenberg-Marquardt (LM) learning function and Tangent Sigmoid as transfer function, which are discussed in the following sections, were used. The ANN was trained using LM technique because it is more powerful and faster than the conventional gradient descent technique (Kisi, 2009 d).

3.1.4.1. Three layered feed forward ANN

A three layered feedforward ANN has an input layer, an output layer, and one hidden layers between the input and output layers. Each of the neurons in a layer is connected to all the neurons of the next layer, and the neurons in one layer are connected only to the neurons of the immediate next layer **(Fig. 3.5)**. The strength of the signal passing from one neuron to the other depends on the weight of the interconnections. The hidden layers enhance the network's ability to model complex functions.

A three-layer feedforward ANN along with a typical processing element is shown in **Fig. 3.5**. Information passes from the input to the output side. The data passing through the connections from one neuron to another are manipulated by weights that control the strength of a passing signal. The nodes in one layer are connected to those in the next, but not to those in the same layer. Thus, the output of a node in a layer is only dependent on the inputs it receives from previous layers and the corresponding weights. The architecture of a typical neuron is also shown in **Fig. 3.5**. Each node multiplies every input by its weight, sums the product, and then passes the sum through a transfer function to produce its result. This transfer function is usually a steadily increasing S-shaped curve, called a sigmoid function. The attenuation at the upper and lower limbs of the ''S'' constrains the raw sums smoothly within fixed limits. The transfer function also introduces a nonlinearity that further enhances the network's ability to model complex functions (Jain and Chalisgaonkar, 2000). The sigmoid function is continuous, differentiable everywhere, and monotonically increasing.



Output = Xo Wo + X1 W1 + …….+ Xn Wn

Fig. 3.5 Typical thee layer feed-forward ANN

3.1.4.2 Training of ANN

In general, it is assumed that the ANN does not have any prior knowledge about the problem. The data enter the network through the input layer. The nodes in the input layer are not computational nodes and simply broadcast the data over weighted connections to the hidden nodes. The ANNs are trained with a set of input and known output pairs called the training set. At the beginning of training the network weights are initialized, either with a set of random values or based on some previous experience. The weights are optimized to get a specific response from an ANN. This process of optimization is called learning/training and is similar to calibration of a hydrological model.

As mentioned earlier, the middle layer neurons take the weighted inputs and sums them. To make a single value output from each neuron, the sum is used in an equation called a transfer function to create an output value. The Levenberg-Marquardt (LM) algorithm, a standard second-order nonlinear least-squares technique, based on the backpropagation process to increase the speed and efficiency of the training was used for training the ANN models.

Suppose X_n represents the input vector, W_{nj} represents the weights between input nodes and hidden nodes, W_{jk} represents the weights between hidden nodes and output nodes, NET_j and NET_k represent the sum of the inputs of the networks, and H_j and Y_k represent the transferred results of a sigmoid function of two networks. The activation value at any node (here j^{th} node), except the input layer nodes, is calculated as

$$NET_j = \sum W_{nj} \times X_n \tag{3.12}$$

This activation value is propagated through an output function at each (here jth) neuron. Activation function could be Sigmoid (Logistic), Hyperbolic tangent (tan-sigmoid, Inverse tangent, threshold, Gaussian radial basis, and Linear, while the first two are the most commonly used in the hydrological modeling (Dawson and Wilby, 2001). In the present study, the hyperbolic tangent sigmoid **(Fig. 3.6)** function was preferred to the logistic function because a multilayer perceptron may learn faster (in terms of the number of epochs required) when the sigmoid function is symmetric than when it is asymmetric (Adeloye and Munari, 2006; Haykin, 1998). Tan-sigmoid transfer function is given by the following formula

$$H_j = f(NET_j) = \frac{e^{(NET_j)} - e^{-(NET_j)}}{e^{(NET_j)} + e^{-(NET_j)}} \tag{3.13}$$

Similarly

$$NET_k = W_{jk} \times H_j \tag{3.14}$$

$$Y_k = f(NET_k) = \frac{e^{(NET_k)} - e^{-(NET_k)}}{e^{(NET_k)} + e^{-(NET_k)}} \tag{3.15}$$

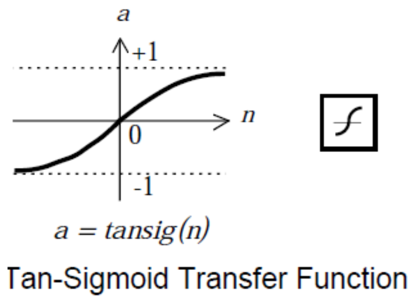


Fig. 3.6. Tan-Sigmoid transfer function

The effect of input is propagated through layers, and the function value for each neuron in the output layer is computed. The sum of the nonlinear least-squares between the observed (target) output, T_k, and computed output, Y_k, defines the error signal, E, that is to be propagated back

$$E^{pt} = \frac{1}{2}\sum_{k=1}^{N} \left(T_k - Y_k\right)^2 \tag{3.16}$$

where E^{pt} = individual pattern error, and N = number of patterns (observations).
If there are m output units, Eq. (3.16) can be written as

$$E^{pt} = \frac{1}{2}\sum_{k=1}^{N}\sum_{p=1}^{m} \left(T_k - Y_k\right)^2 \tag{3.17}$$

The total system error is given by

$$E = \frac{1}{N}\sum_{n=1}^{N} E_n^{pt} \tag{3.18}$$

where E = total system error for all the patterns and is also called a target error. The neural network model stops iteration for training, when E becomes smaller than the target error. This error signal is propagated back, and weights are adjusted to reduce the difference between the desired and computed outputs. In order to find optimal weights (W) and bias (b), training or learning processes must be employed to minimize the error. A number of training algorithms were developed for error back propagation learning. In this study Levenberg-Marquardt (LM) algorithm is used to train the network. LM algorithm has the fasted convergence amongst all algorithms and it would be able to obtain the lowest mean square error in many cases (Cigizoglu and Kisi, 2005; Beale et al. 2012; Lam et al. 2012). LM is a combination of steepest descent and Gaussian-Newton method. Suppose, we have a function *E(W)* which is to be minimized with respect to the parameter vector *W*, then Newton's method would be (Hagan and Menhaj, 1994)

$$\Delta W = -[\nabla^2 E(W)]^{-1} \nabla E(W) \tag{3.19}$$

where $\nabla^2 E(W)$ is the Hessian matrix and ΔE is the gradient. If we assume that *E(W)* is a sum of squares function

$$E(W) = \sum_{i=1}^{N} E_n(W) \tag{3.20}$$

Then it can be shown that

$$\nabla E(W) = J^T(W).E_n(W) \tag{3.21}$$

$$\nabla^2 E(W) = J^T(W)J(W) + S(W) \tag{3.22}$$

where *J(W)*is the Jacobian matrix

$$J(W) = \begin{bmatrix} \frac{\partial E_1}{\partial W_1} & \frac{\partial E_1}{\partial W_2} & \cdots & \frac{\partial E_1}{\partial W_\Omega} \\ \frac{\partial E_2}{\partial W_1} & \frac{\partial E_2}{\partial W_2} & \cdots & \frac{\partial E_2}{\partial W_\Omega} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial E_N}{\partial W_1} & \frac{\partial E_N}{\partial W_2} & \cdots & \frac{\partial E_N}{\partial W_\Omega} \end{bmatrix} \tag{3.23}$$

where Ω and *N* are the number of weights and the number of patterns, respectively. In other words the Jacobian will have as many columns as the number of weights, and the number of rows will be equal to the product of Ω and *N* and

$$S(W) = \sum_{i=1}^{N} E_n(W).\nabla^2 E_n(W) \tag{3.24}$$

According to the Gauss-Newton method, $S(W) \approx 0$, and the one step weight update equation becomes:

$$\Delta W = [J^T(W)J(W)]^{-1} J^T(W)E(W) \tag{3.25}$$

Applying the Levenberg-Marquardt modification to the Gauss-Newton method, the equation becomes:

$$\Delta W = [J^T(W)J(W) + \mu I]^{-1} J^T(W)E(W) \tag{3.26}$$

where I is the identity matrix and $\mu > 0$ which is modified by some factor β in each epoch. For large values of μ, the algorithm becomes steepest descent with the slope of (1/ μ) which is standard backpropagation, and for small values of μ, the algorithm becomes Gauss-Newton method. In the next step, again the network is trained with the new weight:

$$W_{Epoch+1} = W_{Epoch} + \Delta W \tag{3.27}$$

The sum of squares error ($E(W)$) is recomputed using new weights - if $E(W)$ is increased, μ is multiplied by a factor β and if it is reduced μ is divided by β. This procedure will be iterated again and again until the $E(W)$ has been reduced to the error goal, which then the algorithm is assumed to be converged.

This updating of weights is continued until the required level of accuracy is obtained between the target values and computed outputs. After such learning is over, the weights are frozen. A data set, that the ANN has not encountered before, is presented to validate its performance. Depending on the outcome, either the ANN has to be retrained or it can be implemented for its designated use.

3.2 MODEL DEVELOPMENT

3.2.1. Flow Chart

In this research, hybrid models, combining wavelet and artificial neural network (WANN), are developed for Brahmaputra river flow prediction at two stations namely Pandu (u/s) and Pancharatna (d/s) located on Brahmaputra River for the following short term and long term time series with multiple time step lead times:

i) Using daily data: Lead times – 2, 3, 4, 7 and 14 day.

ii) Using weekly data: Lead times – 1 and 2 week.

iii) Using monthly data: Lead time – 1 month.

Here weekly and monthly data means discharge is observed on weekly and monthly basis, respectively. The flow chart for model development is shown **Fig.3.7.**

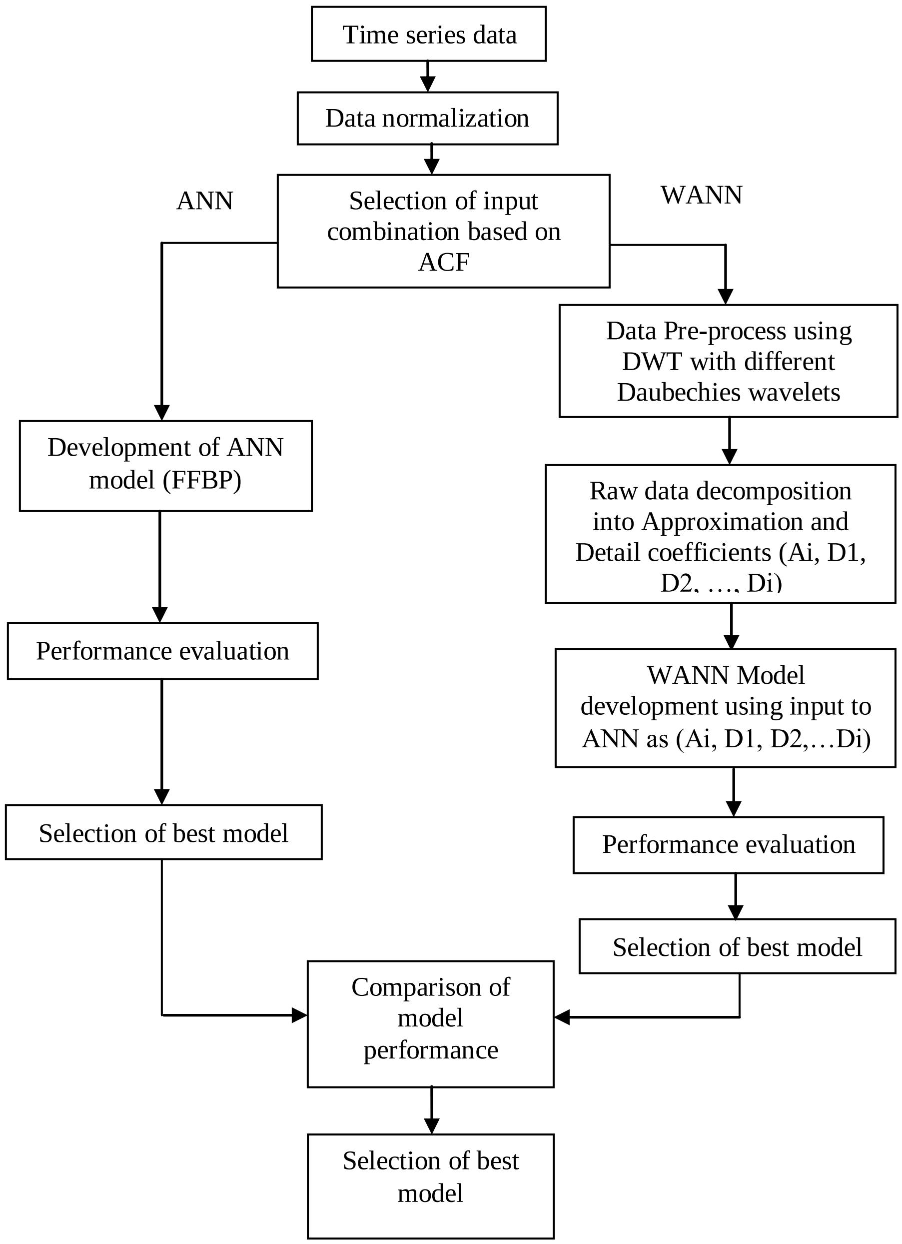


Fig.3.7. Flow chart of model development

3.2.2 Statistical Properties of Observed Flow

To develop models, ten year daily (from Jan 1990 to Dec 1999) flow data at the two stations namely, Pandu (u/s) and Pancharatna (d/s), were collected from Water Resources Department, Assam, India. The observed time series discharge at Pandu and Pancharatna stations are shown in Fig. **3.8 and 3.9,** respectively.

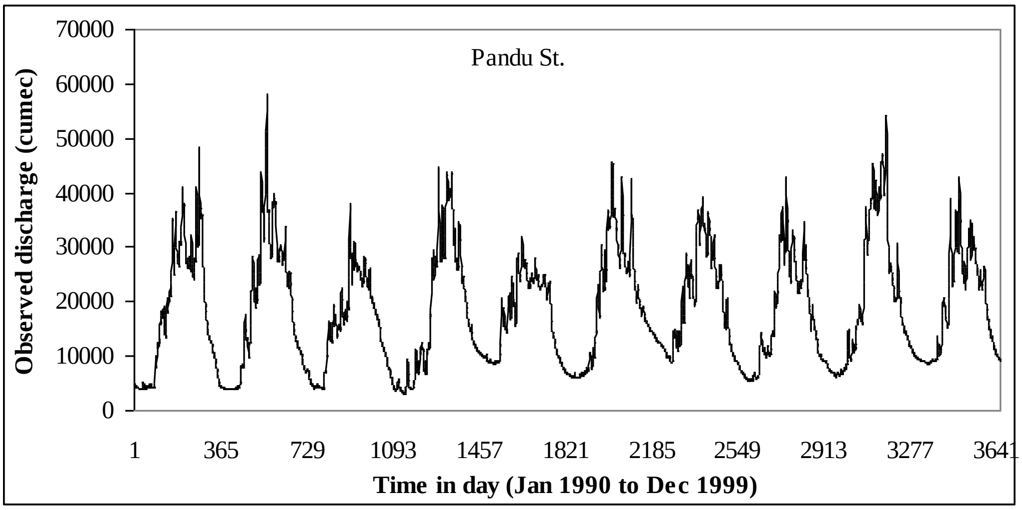


Fig. 3.8. Observed flow series (Pandu Station)

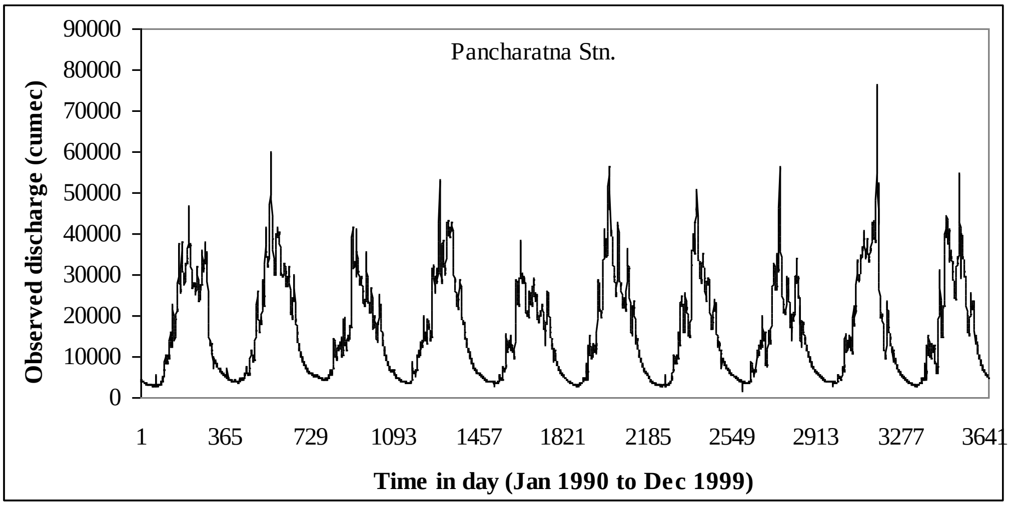


Fig. 3.9. Observed flow series (Pancharatna Station)

From the **Figures 3.8 and 3.9** it is seen that, the discharge is highly non-stationary, especially, during raining season (June to September of every year). In the rising limb (during February to April), also, the discharge is fluctuating because of the contribution of flow due to snow melting.

The statistical characteristics of flow data are shown in the **Table 3.1** below which reveals high variability. In the table Q_{mean}, Q_{max}, Q_{min}, S_d and C_x denotes the mean, maximum, minimum, standard deviation and skewness, respectively.

Table 3.1. Statistical properties of flow data

Statistical Parameter	Pandu Station			Pancharatna Station		
	Training	Testing	All	Training	Testing	All
			Daily data			
Q_{mean} (m^3/s)	17803	19426	18256	16159	16236	16161
Q_{max} (m^3/s)	58200	54100	58200	59832	76236	76236
Q_{min} (m^3/s)	3008	5567	3008	2628	1723	1723
S_d (m^3/s)	10508	10662	10579	11783	12388	11965
C_x	0.507	0.698	0.563	0.726	0.968	0.809
			Weekly data			
Q_{mean} (m^3/s)	17998	18952	18284	16258	16391	16130
Q_{max} (m^3/s)	55800	54100	55800	56244	64284	64284
Q_{min} (m^3/s)	3175	5544	3175	2628	1723	1723
S_d (m^3/s)	10515	10968	10651	11957	12952	11966
C_x	0.495	0.771	0.586	0.786	1.146	0.830
			Monthly data			
Q_{mean} (m^3/s)	18030	18271	18104	15704	15779	15727
Q_{max} (m^3/s)	43900	44200	44200	42901	40728	42901
Q_{min} (m^3/s)	3586	5567	3586	3081	3081	3081
S_d (m^3/s)	10245	10332	10229	11325	11677	11382
C_x	0.429	0.758	0.521	0.557	0.782	0.618

Table 3.1 reveals that the discharge is fluctuating highly in nature at both the stations during study period (Pandu station: minimum = 3008 cumec, maximum = 58200 cumec; Pancharatna station: minimum = 1723 cumec, maximum = 76236 cumec). Standard deviation for Pandu and Pancharatna stations are found to be 10579 cumec and 11965 cumec, respectively, indicating wide dispersion of values from the average one. Also, the observed flows show high positive coefficient of skewness (Pandu: C_x = 0.563; Pancharatna: C_x = 0.809), indicating data has more scattered distribution about mean. To examine the testing capability of the models, two stations with different statistical properties are chosen for study.

3.2.3 Input Selection in ANN and WANN Model

There is no fixed rule for deciding the number of parameters in the input layer of ANN. In any time series forecasting, the values of future time step (Q_{t+n}) is bound to be dependent on antecedent values i.e. Q_t, Q_{t-1}, Q_{t-2}, …, Q_{t-j}. But it is difficult to decide how many lags in past would result in better efficiency i.e. the value of *j* (lag) is not known a priori. Determining *j* plays an important role in hydrologic time series forecasting because this may help in avoiding loss of information that may result if key input variables are omitted, also prevent unnecessary inclusion of input variables which may create problem in training of ANN.

Table 3.2. Correlation coefficients for the flow series (Pandu station)

Output[@] (Q_{t+n})	Input (Q_{t-j})[#]									
	Q_t	Q_{t-1}	Q_{t-2}	Q_{t-3}	Q_{t-4}	Q_{t-5}	Q_{t-6}	Q_{t-7}	Q_{t-8}	Q_{t-9}
Daily data										
Q_{t+2}	0.984	0.972	0.959	0.947	0.935	0.924	0.914	0.905	0.897	0.889
Q_{t+3}	0.972	0.959	0.947	0.935	0.924	0.914	0.905	0.897	0.890	0.884
Q_{t+4}	0.959	0.947	0.935	0.924	0.914	0.905	0.897	0.890	0.884	0.877
Q_{t+7}	0.924	0.914	0.905	0.897	0.890	0.884	0.878	0.872	0.867	0.861
Q_{t+14}	0.872	0.866	0.861	0.855	0.850	0.844	0.838	0.832	0.825	0.816
Weekly data										
Q_{t+1}	0.919	0.863	0.822	0.764	0.703	0.638	0.567	0.485	0.400	
Q_{t+2}	0.863	0.822	0.763	0.702	0.637	0.566	0.483	0.398	0.309	
Monthly data										
Q_{t+1}	0.74	0.43	0.03	-0.38	-0.67	-0.77	-0.68	-0.40		

[@#] n and j are lead time and lag, respectively, in the units of respective time series.

Table 3.3. Correlation coefficients for the flow series (Pancharatna station)

Output[@] (Q_{t+n})	Input (Q_{t-j})[#]									
	Q_t	Q_{t-1}	Q_{t-2}	Q_{t-3}	Q_{t-4}	Q_{t-5}	Q_{t-6}	Q_{t-7}	Q_{t-8}	Q_{t-9}
Daily data										
Q_{t+2}	0.983	0.971	0.957	0.942	0.929	0.915	0.903	0.891	0.880	0.871
Q_{t+3}	0.970	0.957	0.942	0.928	0.915	0.902	0.891	0.880	0.870	0.862
Q_{t+4}	0.957	0.942	0.928	0.915	0.902	0.891	0.880	0.870	0.862	0.854
Q_{t+7}	0.915	0.902	0.891	0.880	0.870	0.862	0.854	0.846	0.839	0.832
Q_{t+14}	0.846	0.839	0.832	0.825	0.818	0.812	0.805	0.799	0.792	0.784
Weekly data										
Q_{t+1}	0.912	0.842	0.793	0.730	0.664	0.601	0.531	0.456	0.372	
Q_{t+2}	0.842	0.793	0.729	0.664	0.600	0.530	0.454	0.371	0.276	
Monthly data										
Q_{t+1}	0.72	0.42	-0.04	-0.42	-0.72	-0.82	-0.72	-0.42		

[@#]n and j are lead time and lag, respectively, in the units of respective time series.

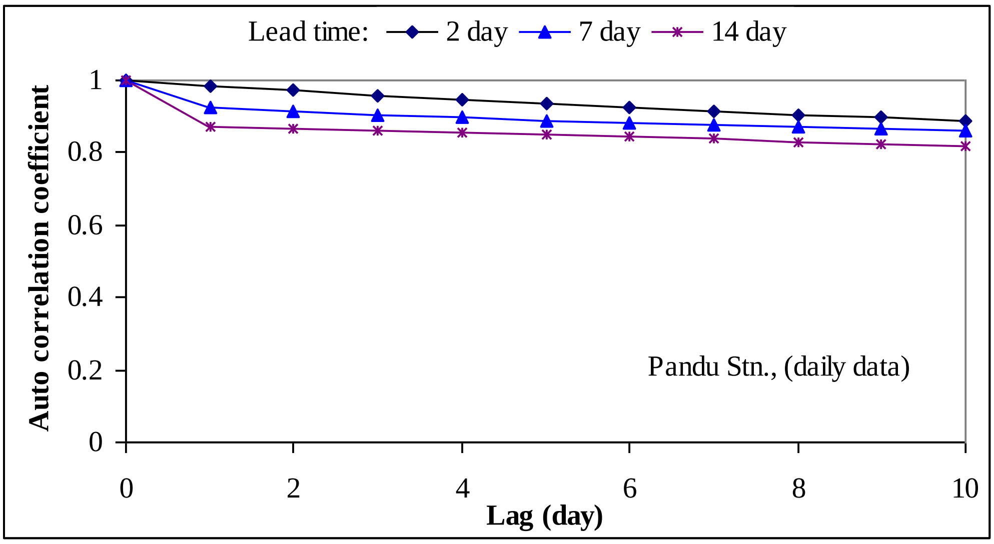


Fig. 3.10. Auto-correlation plot of the flow series for Pandu station (daily data)

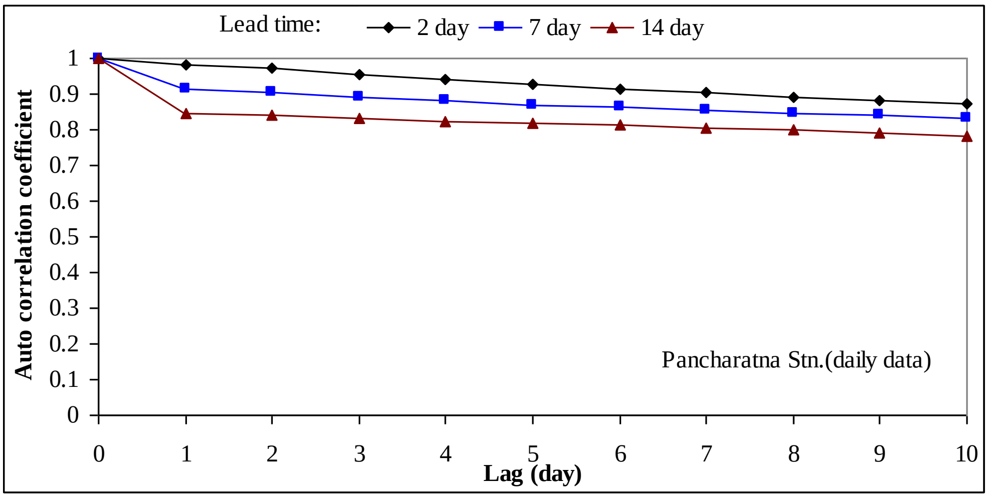


Fig.3.11 Auto-correlation plot of the flow series for Pancharatna Stn. (daily data)

Many researchers (Krishna, 2013; Nayak et al., 2013) have employed the method suggested by Sudheer et al. (2002), which utilizes the statistical properties such as cross-, auto-, and partial-auto-correlation of the data series in identifying a unique input vector that best represents the process for the basin. In the present study the input vectors to models are selected based on the auto-correlation coefficient (shown in **Table 3.2 and 3.3)** between the variables in question. **Figures 3.10 and 3.11** show auto-correlation plots of the flow series for Pandu (daily data) and Pancharatna (daily data) stations, respectively. **Figures 3.10 and 3.11** reveal that, the time series data is highly non-stationary as in non stationary series the auto correlation coefficient do not decay to zero as they decay exponentially to zero in stationary series.

Based on auto correlation, the following combinations containing different numbers of input values were considered for the input layer to predict the value of discharge for different lead times.

1. Q_t;
2. Q_t, $Q_{(t-1)}$;
3. Q_t, $Q_{(t-1)}$, $Q_{(t-2)}$;
4. Q_t, $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$;
5. Q_t, $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$, $Q_{(t-4)}$;
6. Q_t, $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$, $Q_{(t-4)}$, $Q_{(t-5)}$;
7. Q_t, $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$, $Q_{(t-4)}$, $Q_{(t-5)}$, $Q_{(t-6)}$)

After finding input combinations using auto-correlation, the optimal input combinations for every lead time and for every time series is finalized using trial and error procedure by varying number of neurons in hidden layer from 1 to 20 using three layer FFBP network and Lavenberg-Marquardt as training algorithm with tansig as activation function. For the selection of optimal input combinations the total number of trials taken were **1120** {7 (no. of input combinations) × 20 (no. of neurons) × 8 (no. of lead times) for all time series i.e. daily, weekly, and monthly). The optimal combinations, shown in **Table 3.4**, were selected based on lowest RMSE.

Table 3.4 ANN and WANN structure in terms of input and output parameters

Model No.	Time series	Input parameter@	Output parameter# $Q_{(t+n)}$
1	Daily	Q_t, $Q_{(t-1)}$	$Q_{(t+2)}$
2		Q_t, $Q_{(t-1)}$, $Q_{(t-2)}$	$Q_{(t+3)}$
3		Q_t, $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$	$Q_{(t+4)}$
4		Q_t, $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$	$Q_{(t+7)}$
5		Q_t, $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$	$Q_{(t+14)}$
6	Weekly	Q_t, $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$	$Q_{(t+1)}$
7		Q_t, $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$	$Q_{(t+2)}$
8	Monthly	Q_t, $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$	$Q_{(t+1)}$

@ Q_t is current discharge value and $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$ are 1, 2 and 3 time step past discharge values. # n = lead time in the unit of respective time series.

3.2.4 ANN Model Development

Multilayer perceptron (MLP) feed forward backpropagation ANN models without data pre-processing were developed to forecast river discharge for multiple lead times using short term and long term time series data. In most of the hydrologic time series modeling, three layer-feedforward type of artificial neural network is used (Tayfur, 2006). A three-layer (which contains one input layer, one hidden layer, and one output layer) feedforward backpropagation artificial neural network were applied in the study. ANN application for simulation of discharge consists of two stages – training (learning) and testing. The process of training the ANN was carried out with Lavenberg-Marquardt (LM) learning function and Tangent Sigmoid as transfer function for all lead times with respective input combinations mentioned in **Table 3.4**. LM technique was used because it is more powerful and faster than the conventional gradient descent technique (Kisi, 2009 c). After training, the process of testing was carried out. For training and testing first 70 % and last 30 % data, respectively, were used. Each MLP was trained with 1–20 hidden neurons in the hidden layer with Levenberg–Marquardt back propagation as the training algorithm with tansig activation function to optimize the parameters. For all lead times the numbers of neurons in the hidden layer is found by trial-and-error method (Krishna, 2013; Moossavi, 2013). The time series data before going through the network are usually normalized between 0 and 1 (Nourani et al., 2009 a). So the time series flow data during training and testing period is normalized by dividing the discharge value by the maximum discharge. Model with low RMSE in testing period is treated as best model. The optimal learning rate and momentum coefficient used as 0.4 and 0.6, respectively. The termination of iteration was kept at 1000.

3.2.5 WANN Model Development

After developing ANN models for both the stations, hybrid wavelet transform – artificial neural network (WANN) models were developed for the same stations. In WANN models, first of all, the normalized input data were decomposed into approximation and detail coefficients using discrete wavelet transform (DWT). As all hydrological data are observed at discrete time interval, in all WANN models, discrete wavelet transform (DWT) was used for processing of time series data in the form of approximations and details at different levels so that gross and small features of a signal can be separated (Deka and Prahalada, 2012). From the basics of DWT, the maximum level of decomposition is given by the following formula

$$L = 2^M \tag{3.28}$$

$$\text{hence,} \qquad M = [\ln L / \ln 2] \tag{3.29}$$

where L is length of data and M is maximum decomposition level. At largest scale (decomposition level) only one wavelet is required to cover the entire length of signal and only one coefficient is produced). In our study the lengths of data points for daily, weekly and monthly time series during training period are 2557, 365, and 84, respectively. As per the equation 3.29, maximum decomposition level is 11, 8, and 6 (only integer values are taken) for daily, weekly, and monthly data, respectively. However, in our study we have used 7 decomposition levels for daily (2^1 day mode, 2^2 day mode, …, 2^7 day mode) and weekly (2^1 week mode, 2^2 week mode, …, 2^7 week mode) data, while 5 for monthly data (2^1 month mode, 2^2 month mode, …, 2^5 month mode), because the optimum level of decomposition for daily and weekly data was found to lie between 4 and 6, while for monthly 2 and 3, for both the stations. WANN models were developed using the Daubechies function of order 1 to 5 **(Fig.3.13)** and multiresolution level ranging from 1 to 7 for each order of the function (db_il_j, i = 1,…, 5 and j = 1, 2,..,7). In which db refers to the Daubechies function, i is the order of the function, l is the resolution and j is the level of resolution. The choice of mother wavelet depends on the data to be analyzed (Nayak et al. 2013). The shapes of db2 to db5 wavelets are similar to the observed flow series data and that is why these wavelets are used in the present study. However, it was learnt from previous studies that different mother wavelets may be used for improvement of model performance. For a model with j resolution levels there are $j+1$ decomposed time series (one approximation A_j and j detailed i.e. D_1, D_2, ….., D_j). (For example for 4 day lead time the optimum input parameters are Q_t, $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$ and if these inputs are to be decomposed into, say, 4 levels, then there would be 5 decomposed series (A_4, D_1, D_2, D_3, D_4) for each input parameter, and hence total there would be 5 (number of decomposed series) × 4 (number of input parameters) = 20 inputs to ANN). These coefficients of details and approximations were used as input to ANN component of the hybrid model to obtain predicted output. The output signals were kept as original series without decomposition. A multilayer perceptron (MLP) feedforward backpropagation ANN was trained with 2 to 15 neurons in hidden layer using Levenberg-Marquardt training algorithm with tansig as activation function. The schematic diagram of proposed WANN model is shown in **Fig. 3.12.**

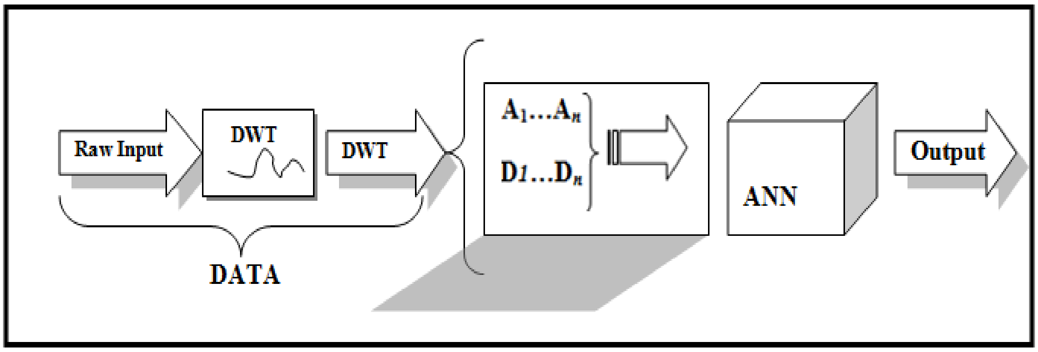


Fig. 3.12. Schematic diagram of the proposed WANN model

3.2.5.1 Selection of Mother Wavelets

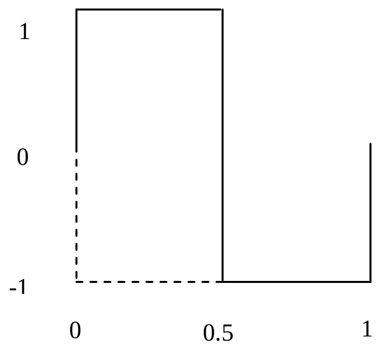
In this study wavelet transforms are used for their denoising capabilities in order to improve ANN forecasting accuracy. Torrence and Compo (1998) described wavelet as a tool to analyze variation of power within a time series. Because of compact support in which wavelets are defined, wavelet filter banks are well suited to decompose, manipulate and represent non stationary time series. Fundamental manuals and practicals guide to wavelet analysis were provoded by Torrence and Compo, 1998; Daubechies, 1992; Mayer, 1992.

In the present study, the focus is on discrete wavelet transform since it produces a useful multiresolution in time frequency analysis compared to continuous wavelet transform which needs a very large memory space with large computational time. One of the most important families of DWT is the orthonormal wavelet transform which is extremely fast algorithm and widely used for analyzing nonstationary data. Orthogonality property provides independency for the detail coefficients and therefore allows the addition of one or more of the detail coefficients of different levels to the approximation coefficient of the first level in different combinations for purposes of various analyses.

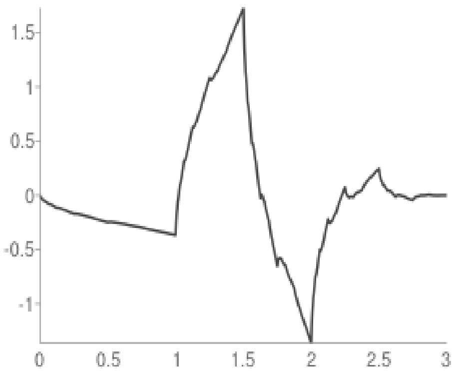
The choice of the mother wavelet depends on the data to be analyzed. In this study, dealing with very irregular signal shape, an irregular wavelet, Daubechies wavelets (Db) of order 1 (db1) to 5 (db5) have been opted. Daubechies wavelets of order 1 to 5 are shown in **Fig. 3.13.** All Daubechies wavelets of order N (db$N$) are asymmetric, orthogonal and biorthogonal. They are compactly supported wavelets with extremal phase and highest number of vanishing moments for a given support width (Misiti, 2010). The support width of Daubechies wavelet is equal to $2N$-1. Daubechies wavelets have no explicit expression except for db1 (haar). The wavelets having compact support or narrow window function are suitable for local analysis of the signal.

The Haar (db1) wavelet is the simple, fast, memory efficient and exactly reversible without edge effects that are problem with other wavelet transforms. The major limitation of Haar wavelet is that the signal denoising is not always so effective, due to the fact that the transform cannot compress the energy of the original signal into a few high energy values lying above the noise threshold. Hence, Haar wavelet transform is not useful in compression and noise removal of signal processing.

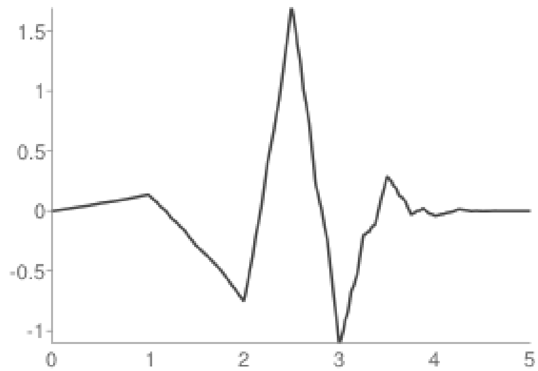
The Daubechies wavelets are a family of orthogonal wavelets defining a discrete wavelet transform and characterized by maximal number of vanishing moments for some given support. With each wavelet type of this class, there is a scaling function (father wavelet) which generates an orthogonal multiresolution analysis. The daubechies wavelets are continuous; thus they are more computationally expensive to use than the Haar wavelet. Daubechies wavelets use overlapping windows, so the high frequency coefficient spectrum reflects all high frequency changes. Therefore Daubechies wavelets are useful in compression and noise removal of signal processing.



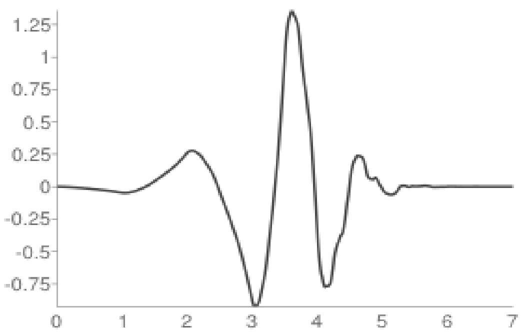
(a) db1



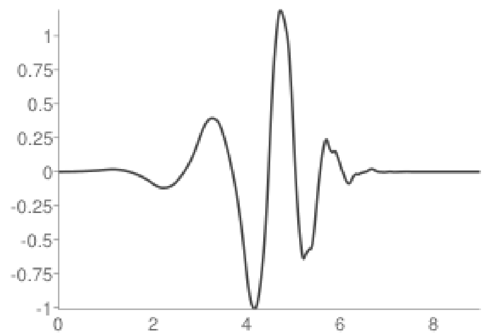
(b) db2



(c) db3



(d) db4



(e) db5

Fig. 3.13. Daubechies wavelets of order 1 to 5

3.2.6 Performance Criteria

Following measures of evaluation have been used to compare the performance of the various models.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \left(Q_{obs} - Q_{com}\right)^2}{N}} \tag{3.30}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N} \left(Q_{obs} - Q_{com}\right)^2}{\sum_{i=1}^{N} \left(Q_{obs} - \bar{Q}_{obs}\right)^2} \tag{3.31}$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} \left| Q_{obs} - Q_{com} \right| \tag{3.32}$$

$$B = \frac{\sum_{i=1}^{N} Q_{comp}}{\sum_{i=1}^{N} Q_{obs}} \tag{3.33}$$

$$SI = \frac{RMSE}{\sum_{i=1}^{N} \bar{Q}_{obs}} \tag{3.34}$$

where *RMSE*, R^2,*MAE*, *B*, *SI*, *N*, Q_{obs}, Q_{com}, and $\bar{Q}_{obs}$ are root-mean squared error (*RMSE*), determination coefficient (R^2), mean absolute error (*MAE*), bias (*B*), scatter index (*SI*), number of observations, observed data, computed values, mean of observed data, respectively. The *RMSE* provides a good measure of goodness of fit at high flows, whereas *MAE* measures a more balanced perspective of the goodness of the fit at moderate flows (Karunanithi et al. 1994). Models with low *RMSE* are treated as best models. R^2 is a correlation measure between computed and observed data. It provides a measure of the ability of a model to predict flows which are different from mean and its value ranges from -∞ at worst case to +1 for a perfect correlation. A R^2 of 0.9 and above is very satisfactory, 0.8 to 0.9 represents a fairly good model, and below 0.8 is deemed unsatisfactory (Dawson and Wilby, 2001). The BIAS (*B*) indicates whether the model is overestimating (B>1) or underestimating (B<1) compared to the observed values. B = 1 indicates non-biased model performance. Scatter index is scalable measure of model precision. The model becomes more precise as the SI reaches zero. (Salvatore et al. 2012).

CHAPTER 4

RESULTS AND DISCUSSION

4.1 INTRODUCTION

In this research, the applicability and generalization capability of wavelet transformation (WT) as a preprocessing technique combined with artificial neural network (ANN) is investigated using case study of forecasting Brahmaputra River flow using daily (short term), weekly (long term), and monthly (long term) time series data for multiple lead times. Ten year daily time series data for the two stations namely Pandu (u/s) and Pancharatna (d/s) located on Brahmaputra River are used in the study. The above two stations, spaced 140 km apart, were selected due to highly varied statistical parameters of observed flow at the stations. (Refer **Table 3.1, pp 39)**. As mentioned earlier in **Chapter 3** the discharge is fluctuating highly in nature at both the stations during study period. To examine the testing capability of the models, two stations with different statistical properties are chosen for study.

Using daily time series data, forecasting is carried out for lead times 2, 3, 4, 7, and 14 day. Using weekly time series data forecasting is carried out for lead times of 1, and 2 week, while using monthly data forecasting is carried out for lead time of 1 month. Total seven input combinations were finalized based on auto-correlations for flow series. Then the optimal input combinations for every lead time and for every time series is finalized using trial and error procedure by varying the number of neurons in hidden layer from 1 to 20 using three layer FFBP network and Lavenberg-Marquardt as training algorithm with tansig as activation function. To forecast 2, and 3 day ahead flow values the optimal input combinations obtained were i) Q_t, $Q_{(t-1)}$, and ii) Q_t, $Q_{(t-1)}$, $Q_{(t-2)}$, respectively. And for all other lead times (for daily, weekly and monthly time series data) the input combinations obtained were Q_t, $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$, where Q_t is current discharge value and $Q_{(t-1)}$, $Q_{(t-2)}$, $Q_{(t-3)}$ are 1, 2 and 3 time step past discharge values.

At the first stage, for Pandu station, a multilayer perceptron (MLP) feed forward backpropagation ANN models without data pre-processing was developed to forecast river discharge for multiple lead times for both short term and long time series data. In most of the hydrologic time series modeling, three layer-feedforward type of artificial neural network is used (Tayfur, 2006). In the present study, backpropagation algorithm with Lavenberg-Marquardt (LM) learning function and Tangent Sigmoid as transfer function were used. The ANN was trained using LM technique because it is more powerful and faster than the conventional gradient descent technique (Kisi, 2009 a). Each MLP was trained with 1–20 hidden neurons in the hidden layer with Levenberg–Marquardt back propagation as the training algorithm with tansig activation function to optimize the parameters. For all lead times the numbers of neurons in the hidden layer is found by trial-and-error method (Krishna, 2013; Moossavi, 2013). The time series data before going through the network are usually normalized between 0 and 1 (Nourani et al., 2009 a). So the time series flow data is normalized by dividing the discharge value by the maximum discharge. The model with low RMSE in testing period is treated as best model.

In the second stage, hybrid wavelet transform – artificial neural network (WANN) models were developed for Pandu station. As all hydrological data are observed at discrete time interval, in all WANN models, discrete wavelet transform (DWT) was used for processing of time series data in the form of approximations and details at different levels.

In the third stage, studies were carried out for Pancharatna station for which data is highly non stationary (high standard deviation and skewness coefficient) compared to Pandu station. Here the same models which have given best results for Pandu station for 2, 3, 4, 7 and 14 day lead times (using daily data), are used to predict flow at Pancharatna station for the same lead times i.e. 2, 3, 4, 7 and 14 day using daily data. While new models by carrying all runs (similar to Pandu station) are developed for lead times of 1 week, 2 week using weekly data and 1 month using monthly data.

For the above applications, out of 10 years data, first 7 years data (Jan 1990 to Dec 1996) are used for training, and remaining 3 years data (Jan 1997 to Dec 1999) for testing. The results of daily, weekly, and monthly scale are discussed in the following sections.

4.2 MODEL RESULTS FOR DAILY TIME SERIES DATA

Optimum results (low RMSE) of model runs (for both ANN & WANN) carried out by varying order of Daubechies wavelet and decomposition level for daily time series data for both the stations are shown in following tables (**Table 4.1 to 4.5**).

Table 4.1 A: Values of statistical parameters for ANN and WANN models Lead time: 2 day (Pandu Station - Daily data)

Model type	Training period					Testing period					Optimum ANN structure
	RMSE	R^2	MAE	BIAS	S.I.	RMSE	R^2	MAE	BIAS	S.I.	
ANN	1636.55	0.976	974.65	1.001	0.092	1636.18	0.976	993.00	0.993	0.084	2-8-1
db1*l*1	1485.20	0.980	849.82	0.999	0.083	1542.21	0.979	879.64	0.991	0.079	4-2-1
db1*l*2	1321.02	0.984	728.16	0.999	0.074	1463.97	0.981	813.70	0.991	0.075	6-2-1
db1*l*3	1222.00	0.986	705.35	0.999	0.068	1297.23	0.985	748.74	0.992	0.067	8-2-1
db1*l*4	1190.93	0.987	684.83	0.999	0.067	1223.86	0.987	715.25	0.996	0.063	10-2-1
db1*l*5	1189.69	0.987	706.26	1.000	0.067	**1220.89**	**0.987**	**733.79**	0.996	**0.063**	**12-2-1**
db1*l*6	1197.51	0.987	713.33	1.000	0.067	1244.62	0.986	740.97	0.994	0.064	14-2-1
db1*l*7	1188.43	0.987	676.74	0.999	0.067	1274.47	0.985	736.45	0.994	0.065	16-2-1
db2*l*1	1455.52	0.981	864.24	1.002	0.082	1537.45	0.979	889.81	0.990	0.079	4-2-1
db2*l*2	1179.96	0.987	654.91	1.001	0.066	1186.49	0.987	639.63	0.993	0.061	6-2-1
db2*l*3	989.61	0.991	563.45	0.999	0.055	1038.13	0.990	566.18	0.995	0.053	8-2-1
db2*l*4	944.40	0.992	549.48	1.001	0.053	**970.41**	**0.992**	**550.04**	**0.998**	**0.049**	**10-2-1**
db2*l*5	963.36	0.991	586.37	1.002	0.054	1016.72	0.991	580.58	0.995	0.052	12-2-1
db2*l*6	952.05	0.992	540.04	1.001	0.053	1045.44	0.992	575.79	0.997	0.054	14-2-1
db2*l*7	988.46	0.991	647.45	1.001	0.055	1093.56	0.989	667.21	0.992	0.056	16-2-1
db3*l*1	1402.81	0.982	854.19	1.001	0.079	1548.26	0.979	877.24	0.989	0.079	4-3-1
db3*l*2	1075.41	0.989	679.74	1.001	0.060	1101.86	0.989	677.97	0.991	0.057	6-3-1
db3*l*3	811.47	0.994	483.83	1.000	0.045	885.96	0.993	521.04	0.997	0.045	8-3-1
db3*l*4	786.87	0.994	495.86	1.003	0.044	**806.45**	**0.994**	**481.36**	**1.001**	**0.041**	**10-3-1**
db3*l*5	781.55	0.994	499.18	0.998	0.044	845.66	0.994	493.39	0.996	0.043	12-3-1
db3*l*6	784.69	0.994	510.26	1.001	0.044	862.10	0.993	503.63	0.998	0.044	14-3-1
db3*l*7	743.62	0.995	486.09	1.002	0.042	930.94	0.992	522.29	0.998	0.048	16-3-1
db4*l*1	1469.43	0.980	870.12	1.000	0.082	1553.88	0.979	925.33	0.992	0.08	4-2-1
db4*l*2	1054.99	0.990	647.89	1.000	0.059	1150.42	0.988	660.90	0.992	0.059	6-2-1
db4*l*3	807.45	0.994	529.92	1.000	0.045	871.39	0.993	549.15	0.994	0.045	8-2-1
db4*l*4	727.89	0.995	507.51	1.002	0.041	842.58	0.994	568.61	0.988	0.043	10-2-1
db4*l*5	723.54	0.995	488.40	1.001	0.040	**779.20**	**0.995**	**509.81**	**0.999**	**0.041**	12-2-1
db4*l*6	722.81	0.995	491.05	1.000	0.040	794.83	0.994	524.36	0.998	0.041	14-2-1
db4*l*7	717.38	0.995	483.57	1.001	0.040	805.61	0.994	536.13	0.997	0.041	16-2-1
db5*l*1	1478.73	0.980	868.50	1.000	0.083	1521.88	0.980	899.95	0.992	0.078	4-2-1
db5*l*2	1126.53	0.988	701.75	1.001	0.063	1105.43	0.989	667.74	0.995	0.057	6-2-1
db5*l*3	816.09	0.994	527.68	0.999	0.046	873.91	0.993	534.26	0.995	0.045	8-2-1
db5*l*4	741.49	0.995	461.91	1.000	0.041	824.55	0.994	477.71	0.997	0.042	10-2-1
db5*l*5	741.35	0.995	473.51	1.001	0.041	**774.72**	**0.995**	**470.21**	**0.998**	**0.039**	**12-2-1**
db5*l*6	739.58	0.995	480.16	1.001	0.041	801.52	0.994	475.64	0.998	0.041	14-2-1
db5*l*7	739.38	0.995	468.72	0.999	0.041	807.13	0.994	473.05	0.998	0.041	16-2-1

Note: RMSE and MAE are in cumec unit.

Table 4.1 B: Values of statistical parameters for ANN and WANN models Lead time: 2 day (Pancharatna station - Daily data)

Model type	Training period					Testing period					Optimum ANN structure
	RMSE	R^2	MAE	BIAS	S.I.	RMSE	R^2	MAE	BIAS	S.I.	
ANN	1764.66	0.977	1044.28	1.002	0.109	2463.33	0.960	1401.83	1.014	0.152	2-8-1
db1*l*5	1199.43	0.989	757.47	1.008	0.074	1795.99	0.979	1144.51	1.018	0.110	12-2-1
db2*l*4	960.30	0.993	540.14	1.001	0.059	1415.36	0.987	796.43	1.007	0.087	10-2-1
db3*l*4	813.84	0.995	469.56	1.001	0.050	1202.33	0.990	698.70	1.007	0.074	10-3-1
db4*l*5	723.40	0.996	440.35	1.002	0.045	1153.34	0.991	672.40	1.006	0.071	12-2-1
db5*l*5	**652.40**	**0.997**	**413.09**	**1.001**	**0.040**	**974.58**	**0.994**	**609.54**	**1.006**	**0.060**	**12-2-1**

Note: RMSE and MAE are in cumec unit.

Table 4.2 A: Values of statistical parameters for ANN and WANN models Lead time: 3 day (Pandu Station - Daily data)

Model type	Training period					Testing period					Optimum ANN structure
	RMSE	R^2	MAE	BIAS	S.I.	RMSE	R^2	MAE	BIAS	S.I.	
ANN	2262.62	0.954	1389.6	0.999	0.127	2254.51	0.955	1423.98	0.991	0.116	3-11-1
db1*l*1	2055.77	0.962	1258.2	0.999	0.115	2082.04	0.962	1294.01	0.986	0.107	6-2-1
db1*l*2	1575.73	0.977	935.84	1.002	0.088	1669.46	0.975	1003.68	0.995	0.086	9-2-1
db1*l*3	1427.93	0.981	875.97	1.001	0.080	1501.37	0.980	917.45	0.995	0.077	12-2-1
db1*l*4	1383.34	0.983	863.32	1.003	0.078	**1493.03**	**0.980**	**913.71**	0.997	**0.076**	**15-2-1**
db1*l*5	1384.38	0.983	852.07	1.001	0.078	1527.15	0.979	929.20	0.996	0.078	18-2-1
db1*l*6	1368.10	0.983	851.20	1.001	0.077	1527.99	0.979	935.53	0.997	0.078	21-2-1
db1*l*7	1394.94	0.982	881.42	1.003	0.078	1551.61	0.979	938.12	0.995	0.079	24-2-1
db2*l*1	2071.29	0.961	1313.5	1.004	0.116	2087.91	0.961	1323.76	0.988	0.107	6-2-1
db2*l*2	1380.19	0.983	817.67	0.997	0.077	1520.80	0.979	944.40	1.010	0.078	9-2-1
db2*l*3	1128.58	0.988	676.23	1.000	0.063	1267.89	0.989	721.82	0.996	0.065	12-2-1
db2*l*4	1097.36	0.989	690.32	1.001	0.061	1230.67	0.987	725.55	0.995	0.063	15-2-1
db2*l*5	1082.54	0.989	658.76	1.001	0.061	**1224.49**	**0.987**	**696.77**	**0.999**	**0.063**	**18-2-1**
db2*l*6	1098.20	0.989	707.63	1.005	0.062	1247.35	0.986	747.51	1.002	0.064	21-2-1
db2*l*7	1099.16	0.989	709.11	1.002	0.062	1262.20	0.986	737.72	0.998	0.065	24-2-1
db3*l*1	2058.30	0.962	1260.5	1.001	0.115	2168.10	0.958	1346.32	0.987	0.111	6-2-1
db3*l*2	1346.97	0.983	836.16	1.004	0.075	1358.82	0.984	888.53	0.998	0.069	9-2-1
db3*l*3	1019.89	0.990	613.18	1.002	0.057	1102.39	0.989	644.53	0.997	0.056	12-2-1
db3*l*4	983.27	0.991	592.15	1.000	0.055	1008.81	0.991	601.56	0.998	0.052	15-2-1
db3*l*5	982.73	0.991	571.09	0.999	0.055	**999.68**	**0.991**	**580.31**	**0.998**	**0.051**	**18-2-1**
db3*l*6	985.13	0.991	574.80	1.000	0.055	1025.59	0.991	591.24	0.998	0.053	21-2-1
db3*l*7	984.53	0.991	575.62	0.999	0.055	1039.87	0.990	596.10	0.995	0.053	24-2-1
db4*l*1	2034.75	0.962	1247.32	1.001	0.114	2167.35	0.958	1372.66	0.989	0.111	6-2-1
db4*l*2	1314.27	0.984	845.16	1.001	0.074	1260.99	0.986	808.31	0.994	0.065	9-2-1
db4*l*3	890.09	0.993	573.58	1.000	0.050	952.42	0.992	577.72	0.998	0.049	12-2-1
db4*l*4	860.35	0.993	559.78	0.999	0.048	937.84	0.992	552.77	0.997	0.048	15-2-1
db4*l*5	872.36	0.993	573.29	1.001	0.049	**925.38**	**0.992**	**557.69**	**0.998**	**0.047**	**18-2-1**
db4*l*6	855.84	0.993	549.61	1.000	0.048	928.20	0.992	548.72	0.998	0.048	21-2-1
db4*l*7	854.76	0.993	547.20	0.999	0.048	950.72	0.992	548.10	0.997	0.049	24-2-1
db5*l*1	2031.40	0.963	1226.82	0.998	0.114	2109.51	0.961	1312.74	0.990	0.108	6-3-1
db5*l*2	1113.05	0.989	731.44	1.001	0.062	1137.23	0.989	736.44	0.995	0.058	9-3-1
db5*l*3	756.94	0.995	463.44	0.999	0.042	848.19	0.994	514.54	1.000	0.044	12-3-1
db5*l*4	729.57	0.995	452.10	0.999	0.041	791.78	0.994	474.78	0.999	0.041	15-3-1
db5*l*5	727.98	0.995	446.72	0.999	0.041	787.11	0.994	472.75	0.999	0.040	18-3-1
db5*l*6	735.84	0.995	452.62	0.999	0.041	**784.13**	**0.995**	**470.08**	**0.999**	**0.040**	**21-3-1**
db5*l*7	783.60	0.994	527.88	1.001	0.044	824.68	0.994	516.47	0.997	0.042	24-3-1

Note: RMSE and MAE are in cumec unit.

Table 4.2 B: Values of statistical parameters for ANN and WANN models Lead time: 3 day (Pancharatna station - Daily data)

Model type	Training period					Testing period					Optimum ANN structure
	RMSE	R^2	MAE	BIAS	S.I.	RMSE	R^2	MAE	BIAS	S.I.	
ANN	2422.19	0.958	1491.4	1.003	0.149	3157.79	0.935	1859.03	1.018	0.194	3-11-1
db1*l*4	1520.95	0.983	906.48	0.998	0.094	2239.51	0.967	1357.38	1.005	0.138	15-2-1
db2*l*5	1195.08	0.989	700.71	1.001	0.074	1702.97	0.981	983.77	1.006	0.105	18-2-1
db3*l*5	1014.17	0.992	649.58	1.006	0.062	1371.69	0.988	906.18	1.016	0.084	18-2-1
db4*l*5	865.09	0.995	536.88	0.999	0.053	1356.89	0.988	782.60	1.005	0.083	18-2-1
db5*l*6	764.10	0.996	496.09	0.999	0.047	**1144.08**	**0.991**	**720.30**	**1.006**	**0.070**	**21-3-1**

Note: RMSE and MAE are in cumec unit.

Table 4.3 A: Values of statistical parameters for ANN and WANN models Lead time: 4 day (Pandu Station - Daily data)

Model	Training period					Testing period					Optimum
type	RMSE	R^2	MAE	BIAS	S.I.	RMSE	R^2	MAE	BIAS	S.I.	ANN structure
ANN	2743.17	0.932	1773.40	1.002	0.154	2727.18	0.934	1792.17	0.988	0.140	4-19-1
db1*l*1	2544.66	0.941	1641.70	1.001	0.143	2931.22	0.924	1778.92	0.993	0.150	8-3-1
db1*l*2	2273.00	0.953	1409.30	1.002	0.123	2299.02	0.953	1415.33	0.988	0.118	12-3-1
db1*l*3	1914.25	0.967	1091.50	0.999	0.107	1957.61	0.966	1129.12	0.994	0.100	16-3-1
db1*l*4	1881.86	0.968	1184.30	1.006	0.105	1905.49	0.968	1164.51	0.992	0.098	20-3-1
db1*l*5	1821.45	0.970	1082.70	1.000	0.102	**1846.17**	**0.970**	**1099.22**	**0.998**	**0.095**	**24-3-1**
db1*l*6	1854.61	0.967	1126.80	1.002	0.104	1938.38	0.967	1180.88	0.998	0.099	28-3-1
db1*l*7	1963.98	0.965	1305.60	1.005	0.110	2214.79	0.957	1383.03	0.988	0.114	32-3-1
db2*l*1	2578.67	0.939	1616.30	0.999	0.145	2700.99	0.935	1701.41	0.981	0.138	8-2-1
db2*l*2	1937.04	0.966	1226.30	0.999	0.109	1956.79	0.966	1215.68	0.987	0.100	12-2-1
db2*l*3	1657.63	0.975	1155.90	1.009	0.093	1855.42	0.969	1153.46	0.983	0.095	16-2-1
db2*l*4	1291.09	0.985	802.63	0.999	0.072	1305.94	0.985	781.37	0.995	0.067	20-2-1
db2*l*5	1297.26	0.985	839.71	1.004	0.073	**1305.12**	**0.985**	**788.04**	**0.999**	**0.067**	**24-2-1**
db2*l*6	1289.24	0.985	848.63	1.001	0.072	1358.15	0.984	827.26	0.998	0.069	28-2-1
db2*l*7	1300.12	0.985	865.85	1.003	0.073	1445.81	0.981	900.14	0.996	0.074	32-2-1
db3*l*1	2594.78	0.939	1629.40	0.999	0.145	2667.94	0.937	1716.67	0.982	0.137	8-2-1
db3*l*2	1656.63	0.975	1045.90	1.001	0.093	1832.61	0.970	1147.39	0.990	0.094	12-2-1
db3*l*3	1096.44	0.989	723.51	1.003	0.061	1198.73	0.987	758.64	0.998	0.061	16-2-1
db3*l*4	1027.58	0.990	681.92	0.999	0.057	1192.68	0.987	750.27	0.994	0.061	20-2-1
db3*l*5	1011.27	0.991	668.16	1.001	0.057	**1136.34**	**0.988**	**718.48**	**0.998**	**0.058**	**24-2-1**
db3*l*6	1013.04	0.991	663.26	1.001	0.057	1162.38	0.988	718.28	0.997	0.059	28-2-1
db3*l*7	1027.41	0.990	640.93	0.999	0.057	1165.10	0.988	698.99	0.995	0.059	32-2-1
db4*l*1	2622.39	0.937	1712.62	1.009	0.147	2660.76	0.937	1764.44	0.988	0.136	8-2-1
db4*l*2	1684.13	0.974	1037.49	0.999	0.094	1705.27	0.974	1062.52	0.989	0.087	12-2-1
db4*l*3	1048.64	0.990	685.48	1.002	0.059	1075.87	0.989	661.02	0.997	0.055	16-2-1
db4*l*4	987.43	0.991	618.57	1.001	0.055	1029.18	0.991	608.59	0.998	0.053	20-2-1
db4*l*5	984.67	0.991	637.70	0.999	0.055	**1003.16**	**0.991**	**626.22**	**0.998**	**0.051**	**24-2-1**
db4*l*6	990.65	0.991	647.20	1.002	0.055	1034.08	0.990	655.49	1.000	0.053	28-2-1
db4*l*7	1000.96	0.991	635.14	1.001	0.056	1054.22	0.990	623.69	0.997	0.054	32-2-1
db5*l*1	2596.07	0.939	1619.85	1.003	0.146	2586.61	0.941	1659.72	0.987	0.133	8-2-1
db5*l*2	1640.08	0.975	1014.84	1.001	0.092	1661.74	0.975	1029.26	0.992	0.085	12-2-1
db5*l*3	980.04	0.991	619.74	1.001	0.055	999.23	0.991	621.25	0.997	0.051	16-2-1
db5*l*4	935.45	0.992	605.48	1.006	0.052	938.45	0.992	587.69	1.000	0.048	20-2-1
db5*l*5	914.17	0.992	567.40	1.001	0.051	**911.97**	**0.993**	**557.09**	**0.999**	**0.046**	**24-2-1**
db5*l*6	907.57	0.992	558.07	1.001	0.051	942.25	0.992	569.61	0.999	0.048	28-2-1
db5*l*7	918.04	0.992	546.53	0.999	0.051	978.32	0.991	587.81	0.996	0.050	32-2-1

Note: RMSE and MAE are in cumec unit.

Table 4.3 B: Values of statistical parameters for ANN and WANN models Lead time: 4 day (Pancharatna station - Daily data)

Model	Training period					Testing period					Optimum
type	RMSE	R^2	MAE	BIAS	S.I.	RMSE	R^2	MAE	BIAS	S.I.	ANN structure
ANN	3133.09	0.929	2067.41	1.005	0.194	3890.35	0.901	2566.07	1.031	0.239	4-19-1
db1*l*5	1918.22	0.973	1229.16	1.008	0.118	2590.84	0.956	1717.58	1.018	0.159	24-3-1
db2*l*5	1431.04	0.985	868.66	1.001	0.088	1986.38	0.974	1137.20	1.009	0.122	24-2-1
db3*l*5	1280.06	0.988	790.25	1.001	0.079	1683.59	0.982	1034.97	1.004	0.104	24-2-1
db4*l*5	1082.75	0.991	691.25	1.004	0.067	1579.20	0.984	982.56	1.012	0.097	24-2-1
db5*l*5	965.52	0.993	612.51	1.003	0.059	**1335.98**	**0.988**	**823.89**	**1.008**	**0.082**	**24-2-1**

Note: RMSE and MAE are in cumec unit.

Table 4.4 A: Values of statistical parameters for ANN and WANN models Lead time: 7 day (Pandu Station - Daily data)

Model	Training period					Testing period					Optimum
type	RMSE	R^2	MAE	BIAS	S.I.	RMSE	R^2	MAE	BIAS	S.I.	ANN structure
ANN	3853.05	0.865	2618.31	1.001	0.212	3834.86	0.872	2481.24	0.971	0.207	4-3-1
db1*l*1	3760.31	0.812	2553.92	1.001	0.206	3765.29	0.876	2410.94	0.969	0.203	8-2-1
db1*l*2	3475.55	0.890	2355.15	1.004	0.191	3497.16	0.893	2278.59	0.976	0.189	12-2-1
db1*l*3	2981.96	0.919	1934.73	1.008	0.164	2989.86	0.922	1762.10	0.978	0.161	16-2-1
db1*l*4	2560.24	0.940	1645.00	1.005	0.141	2635.71	0.939	1578.13	0.986	0.142	20-2-1
db1*l*5	2538.58	0.941	1619.22	1.001	0.139	2487.56	0.946	1482.93	0.991	0.134	24-2-1
db1*l*6	2422.39	0.945	1557.51	1.010	0.133	**2468.65**	**0.947**	**1491.13**	**1.002**	**0.133**	**28-2-1**
db1*l*7	2624.39	0.937	1808.85	1.007	0.144	2662.73	0.938	1643.74	0.984	0.144	32-2-1
db2*l*1	3752.16	0.872	2559.08	1.005	0.206	3743.96	0.878	2475.67	0.979	0.202	8-2-1
db2*l*2	3518.24	0.888	2317.81	1.003	0.193	3422.78	0.898	2261.76	0.986	0.185	12-2-1
db2*l*3	2537.32	0.942	1665.39	1.008	0.139	2544.47	0.943	1560.36	0.979	0.137	16-2-1
db2*l*4	2173.03	0.957	1487.58	1.012	0.119	2245.68	0.956	1318.36	0.984	0.121	20-2-1
db2*l*5	1896.09	0.967	1188.29	1.004	0.104	**1659.89**	**0.976**	**1002.77**	**0.997**	**0.0897**	**24-2-1**
db2*l*6	1850.22	0.969	1242.94	1.002	0.101	1862.44	0.969	1159.95	0.992	0.1006	28-2-1
db2*l*7	1964.51	0.965	1229.79	0.998	0.108	1871.77	0.969	1133.32	0.986	0.101	32-2-1
db3*l*1	3739.91	0.873	2514.77	1.004	0.205	3761.41	0.877	2518.45	0.983	0.203	8-2-1
db3*l*2	3435.34	0.893	2319.89	1.006	0.189	3420.30	0.898	2194.52	0.979	0.175	12-2-1
db3*l*3	2106.51	0.958	1444.77	0.999	0.115	2180.25	0.958	1400.14	0.991	0.118	16-2-1
db3*l*4	1593.15	0.977	1072.26	1.000	0.087	1691.33	0.975	1070.79	0.996	0.091	20-2-1
db3*l*5	1498.24	0.979	980.02	0.999	0.082	1629.66	0.977	990.70	0.994	0.088	24-2-1
db3*l*6	1494.88	0.979	984.06	1.002	0.082	**1623.20**	**0.977**	**995.99**	**0.997**	**0.088**	**28-2-1**
db3*l*7	1547.07	0.978	1069.14	1.000	0.085	1706.08	0.975	1111.52	0.992	0.092	32-2-1
db4*l*1	3780.54	0.870	2589.45	1.006	0.207	3766.10	0.876	2429.88	0.973	0.203	8-2-1
db4*l*2	3378.23	0.896	2300.12	1.005	0.185	3307.23	0.905	2101.06	0.976	0.178	12-2-1
db4*l*3	1930.42	0.966	1259.56	1.003	0.106	2011.94	0.965	1297.45	0.992	0.109	16-2-1
db4*l*4	1503.90	0.979	959.52	1.000	0.083	1524.60	0.979	932.09	0.996	0.082	20-2-1
db4*l*5	1461.88	0.981	950.01	1.001	0.080	**1480.65**	**0.981**	**913.54**	**0.995**	**0.080**	**24-2-1**
db4*l*6	1478.74	0.980	932.35	1.000	0.081	1526.47	0.979	918.39	0.994	0.082	28-2-1
db4*l*7	1393.70	0.982	934.99	1.001	0.076	1717.46	0.974	1076.92	1.000	0.093	32-2-1
db5*l*1	3770.14	0.871	2638.85	1.015	0.207	3715.21	0.879	2504.05	0.989	0.201	8-2-1
db5*l*2	3325.56	0.899	2218.80	1.003	0.182	3277.65	0.906	2057.40	0.975	0.177	12-2-1
db5*l*3	1690.98	0.974	1165.66	1.002	0.093	1743.09	0.973	1090.56	0.993	0.094	16-2-1
db5*l*4	1291.66	0.985	848.31	1.002	0.071	1262.39	0.986	788.16	0.996	0.068	20-2-1
db5*l*5	1239.79	0.986	806.40	1.000	0.068	**1248.16**	**0.986**	**768.34**	**0.996**	**0.067**	**24-2-1**
db5*l*6	1221.19	0.986	791.69	1.001	0.067	1249.34	0.986	770.06	0.997	0.067	28-2-1
db5*l*7	1246.60	0.986	807.52	0.999	0.068	1271.73	0.986	798.13	0.994	0.069	32-2-1

Note: RMSE and MAE are in cumec unit.

Table 4.4 B: Values of statistical parameters for ANN and WANN models Lead time: 7 day (Pancharatna station - Daily data)

Model	Training period					Testing period					Optimum
type	RMSE	R^2	MAE	BIAS	S.I.	RMSE	R^2	MAE	BIAS	S.I.	ANN structure
ANN	4308.72	0.866	2801.05	1.002	0.266	5185.33	0.825	3441.99	1.032	0.318	4-3-1
db1*l*6	2830.19	0.942	1961.30	1.004	0.175	4131.69	0.889	2942.38	1.016	0.254	28-2-1
db2*l*5	2038.56	0.970	1408.02	1.004	0.126	2894.02	0.945	1927.13	1.024	0.178	24-2-1
db3*l*6	1683.52	0.979	1151.42	1.003	0.104	2525.32	0.958	1639.74	1.013	0.155	28-2-1
db4*l*5	1430.24	0.985	918.64	1.001	0.088	2045.91	0.973	1281.54	1.008	0.126	24-2-1
db5*l*5	1411.79	0.986	896.52	1.001	0.087	**1771.43**	**0.979**	**1127.24**	**1.007**	**0.109**	**24-2-1**

Note: RMSE and MAE are in cumec unit.

Table 4.5 A: Values of statistical parameters for ANN and WANN models Lead time: 14 day (Pandu Station - Daily data)

Model	**Training period**					**Testing period**					**Optimum**
type	**RMSE**	**R^2**	**MAE**	**BIAS**	**S.I.**	**RMSE**	**R^2**	**MAE**	**BIAS**	**S.I.**	**ANN structure**
ANN	4907.49	0.780	3576.03	1.002	0.275	5488.50	0.732	3947.75	0.954	0.279	4-1-1
db1*l*1	4817.71	0.789	3441.27	0.992	0.270	5823.80	0.699	4045.09	0.948	0.297	8-8-1
db1*l*2	4812.20	0.789	3545.42	1.020	0.270	5286.64	0.752	3816.84	0.975	0.269	12-8-1
db1*l*3	4507.15	0.815	3130.46	0.999	0.253	5204.55	0.759	3620.07	0.963	0.265	16-8-1
db1*l*4	3865.85	0.864	2603.53	0.999	0.217	5100.61	0.769	3329.82	0.959	0.260	20-8-1
db1*l*5	3168.86	0.909	2102.05	0.998	0.178	4106.67	0.85	2765.57	0.978	0.209	24-8-1
db1*l*6	2771.41	0.93	1962.28	1.007	0.155	**3720.03**	**0.877**	**2485.80**	**0.991**	**0.189**	**28-8-1**
db1*l*7	2706.19	0.933	1887.95	1.002	0.152	4013.96	0.857	2701.82	0.978	0.204	32-8-1
db2*l*1	4885.56	0.783	3564.96	1.008	0.274	5458.75	0.735	3936.44	0.964	0.278	8-3-1
db2*l*2	4816.79	0.789	3454.82	0.994	0.270	5457.22	0.736	3872.70	0.949	0.278	12-3-1
db2*l*3	4543.95	0.812	3085.65	0.979	0.255	5432.43	0.738	3631.44	0.936	0.277	16-3-1
db2*l*4	3189.25	0.907	2048.57	0.999	0.179	4284.64	0.837	2679.13	0.981	0.218	20-3-1
db2*l*5	2700.42	0.995	1860.77	1.004	0.151	3291.19	0.904	2071.54	0.966	0.166	24-3-1
db2*l*6	2619.77	0.938	1831.58	1.004	0.147	**2919.68**	**0.924**	**1924.33**	**0.976**	**0.149**	**28-3-1**
db2*l*7	2634.58	0.937	1804.53	0.999	0.148	2950.23	0.923	1962.05	0.971	0.150	32-3-1
db3*l*1	4885.59	0.783	3545.71	1.001	0.274	5547.27	0.727	3937.69	0.948	0.283	8-2-1
db3*l*2	4795.08	0.791	3536.24	1.009	0.269	5443.48	0.737	3894.96	0.957	0.277	12-2-1
db3*l*3	4485.24	0.817	3113.22	1.003	0.252	5162.39	0.763	3547.50	0.955	0.263	16-2-1
db3*l*4	3233.59	0.905	2258.12	0.999	0.181	3916.88	0.864	2508.19	0.964	0.199	20-2-1
db3*l*5	2428.78	0.946	1656.16	1.003	0.136	2703.44	0.935	1810.59	**0.982**	**0.138**	24-2-1
db3*l*6	2228.95	0.955	1452.60	0.999	0.125	2480.17	0.945	1636.81	0.993	0.126	28-2-1
db3*l*7	2237.72	0.954	1462.82	1.006	0.126	**2434.86**	**0.947**	**1642.17**	**1.001**	**0.124**	**32-2-1**
db4*l*1	4867.54	0.785	3533.37	1.001	0.273	5488.52	0.732	3927.24	0.953	0.279	8-2-1
db4*l*2	4756.76	0.794	3436.21	1.005	0.267	5314.93	0.749	3806.54	0.959	0.271	12-2-1
db4*l*3	4461.18	0.819	3127.51	1.005	0.250	5123.90	0.767	3543.09	0.958	0.261	16-2-1
db4*l*4	3027.74	0.917	2087.62	1.007	0.169	3549.75	0.888	2355.28	0.973	0.181	20-2-1
db4*l*5	2312.40	0.951	1577.28	0.999	0.129	2418.94	0.948	1582.84	0.978	0.123	24-2-1
db4*l*6	2173.61	0.957	1437.45	1.003	0.122	**2321.45**	**0.952**	**1524.53**	**0.986**	**0.118**	**28-2-1**
db4*l*7	2237.51	0.954	1553.02	1.007	0.125	2448.96	0.947	1586.02	0.983	0.125	32-2-1
db5*l*1	4871.26	0.784	3571.83	1.006	0.273	5587.47	0.723	4108.05	0.965	0.285	8-2-1
db5*l*2	4733.27	0.796	3398.59	1.001	0.266	5288.24	0.752	3764.32	0.957	0.269	12-2-1
db5*l*3	4479.21	0.818	3132.75	0.998	0.251	4634.11	0.809	3300.98	0.968	0.236	16-2-1
db5*l*4	2780.46	0.929	1977.46	1.008	0.156	3213.09	0.908	2160.44	0.975	0.164	20-2-1
db5*l*5	1962.85	0.965	1326.28	0.997	0.110	2401.44	0.949	1545.91	0.991	0.122	24-2-1
db5*l*6	1946.40	0.965	1261.87	0.999	0.109	**<u>2228.35</u>**	**<u>0.956</u>**	**<u>1433.47</u>**	**<u>0.997</u>**	**<u>0.113</u>**	**28-2-1**
db5*l*7	1961.61	0.965	1305.11	1.001	0.110	2246.58	0.955	1463.31	0.992	0.114	32-2-1

Note: RMSE and MAE are in cumec unit.

Table 4.5 B: Values of statistical parameters for ANN and WANN models Lead time: 14 day (Pancharatna station - Daily data)

Model	**Training period**					**Testing period**					**Optimum**
type	**RMSE**	**R^2**	**MAE**	**BIAS**	**S.I.**	**RMSE**	**R^2**	**MAE**	**BIAS**	**S.I.**	**ANN structure**
ANN	5704.97	0.765	4035.67	1.014	0.353	7084.21	0.673	5116.37	1.068	0.433	4-1-1
db1*l*6	4173.09	0.874	2765.27	1.007	0.258	4913.55	0.843	3425.92	1.036	0.300	28-8-1
db2*l*6	3071.96	0.932	2062.78	1.004	0.190	4716.30	0.855	3017.38	1.038	0.288	28-3-1
db3*l*7	2420.95	0.958	1737.07	1.014	0.149	3848.56	0.904	2663.88	1.031	0.235	32-2-1
db4*l*6	2254.95	0.963	1450.11	0.997	0.139	3328.27	0.928	2173.63	1.001	0.203	28-2-1
db5*l*6	2100.07	0.968	1423.29	1.002	0.129	**<u>3100.53</u>**	**<u>0.937</u>**	**<u>2027.85</u>**	**<u>1.013</u>**	**<u>0.189</u>**	**28-2-1**

Note: RMSE and MAE are in cumec unit.

4.2.1 ANN Model Results

It can be seen from **Table 4.1 – 4.5** that for ANN model for Pandu station in testing period the values of determination coefficient (R^2), root mean squared error (RMSE), mean absolute error (MAE), changes with respect to lead time forecast. The R^2 values were found to decrease from 0.976 for 2 day lead time to 0.732 for 14 day lead time. The RMSE increases from 1636.18 cumec for 2 day lead time to 5488.50 cumec for 14 day lead time. Also, MAE values increase from 993.00 to 3947.75 cumec for 2 day and 14 day lead times, respectively. BIAS values are decreasing with lead time (2 day lead time: 0.993; 14 day lead time: 0.954), indicating models underestimation capacity is increasing. Also, scatter index is increasing from 0.084 for 2 day lead time to 0.279 for 14 day lead time, indicating models precision is decreasing. While, during training period, the R^2 values were found to decrease from 0.976 for 2 day lead time to 0.780 for 14 day lead time. RMSE increases from 1636.55 cumec for 2 day lead time to 4907.49 cumec for 14 day lead time. Also, MAE values increased from 974.65 to 3576.03 cumec for 2 day and 14 day lead times, respectively. BIAS values are very close to 1 for all lead times, indicating model is unbiased. Also, scatter index is increasing from 0.092 for 2 day lead time to 0.275 for 14 day lead time, indicating models precision is decreasing. Similar trend is observed for Pancharatna station. In comparison with Pancharatna, Pandu station results are better.

It is found that the model efficiency is decreasing with increase in lead time. This may be due to significant fluctuations of the data around mean values such as high standard deviation (**Table 3.1, pp 39**). **Tables 4.1 – 4.5** also show optimum ANN structure (e.g. for 3 day lead time, the meaning of 3-11-1 is that, 3 neurons in input layer, 11 neurons in hidden layer and 1 neuron in output layer).

4.2.2 WANN Model Results

The normalized observed data was decomposed using Daubechies wavelets of order 1 (db1) to 5 (db5) upto 7th level decomposition, which were fed as input to ANN, making the model as WANN(db$_i$), where *i* is the order of Daubechies wavelet. **Figs. 4.1** (**b – f**) show approximation coefficients obtained using db1 to db5 wavelets along with observed flow data (**Fig. 4.1 a**) for Pandu station (daily data) during training period. **Fig. 4.1** reveals that with increase in wavelet order the approximation signal becomes more and smoother (stationary). **Fig. 4.2** shows the details and approximation signals at decomposition level 5 using db5 as mother wavelet for the whole daily data of Pandu station. From **Fig. 4.2** it is clear that, high frequency components of the original time series are captured in the first resolution level and with increase in decomposition level (scale) signal becomes more and more smoother (stationary). The number of neurons in hidden layer was computed by trial-and-error method and performances of these are tabulated in **Table 4.1 – 4.5.** Results of **Table 4.1 – 4.5** reveal that the performance of wavelet based hybrid WANN model is much better than the regular ANN model in testing period.

For Pandu station (during training period), from the analysis it was found that the value of R^2 decreased from 0.995 (db5*l*5 – WANN(db5)) for 2 day lead time to 0.965 (db5*l*6 – WANN(db5)) for 14 day lead time, while for ANN model this decrease was from 0.976 to 0.780. Also, the RMSE values increased from 741.35 (db5*l*5) to 1946.40 (db5*l*6) cumec for 2 day and 14 day lead time, respectively, while for ANN model this increase was from 1636.55 to 4907.49 cumec. The MAE values were found to increase from 473.51 (db5*l*5) to 1261.87 (db5*l*6) cumec for 2 day and 14 day lead time, respectively, while for ANN model this increase was from 974.65 to 3576.03 cumec. The scatter index was found to increase from 0.041 (db5*l*5) to 0.109 (db5*l*6) for 2 day and 14 day lead time, respectively, while for ANN model this increase was from 0.092 to 0.275. While BIAS values are found to very close to 1.For Pandu station (during testing period), from the analysis it was found that the value of R^2 decreased from 0.995 (db5*l*5 – WANN(db5)) for 2 day lead time to 0.956 (db5*l*6 – WANN(db5)) for 14 day lead time, while for ANN model this decrease was from 0.976 to 0.732. Also, the RMSE values increased from 774.72 (db5*l*5) to 2228.35 (db5*l*6) cumec for 2 day and 14 day lead time, respectively, while for ANN model this increase was from 1636.18 to 5488.50 cumec. The MAE values were found to increase from 470.21 (db5*l*5) to 1433.47 (db5*l*6) cumec for 2 day and 14 day lead time, respectively, while for ANN model this increase was from 993.00 to 3947.75 cumec. The scatter index was found to increase from 0.039 (db5*l*5) to 0.113 (db5*l*6) for 2 day and 14 day lead time, respectively, while for ANN model this increase was

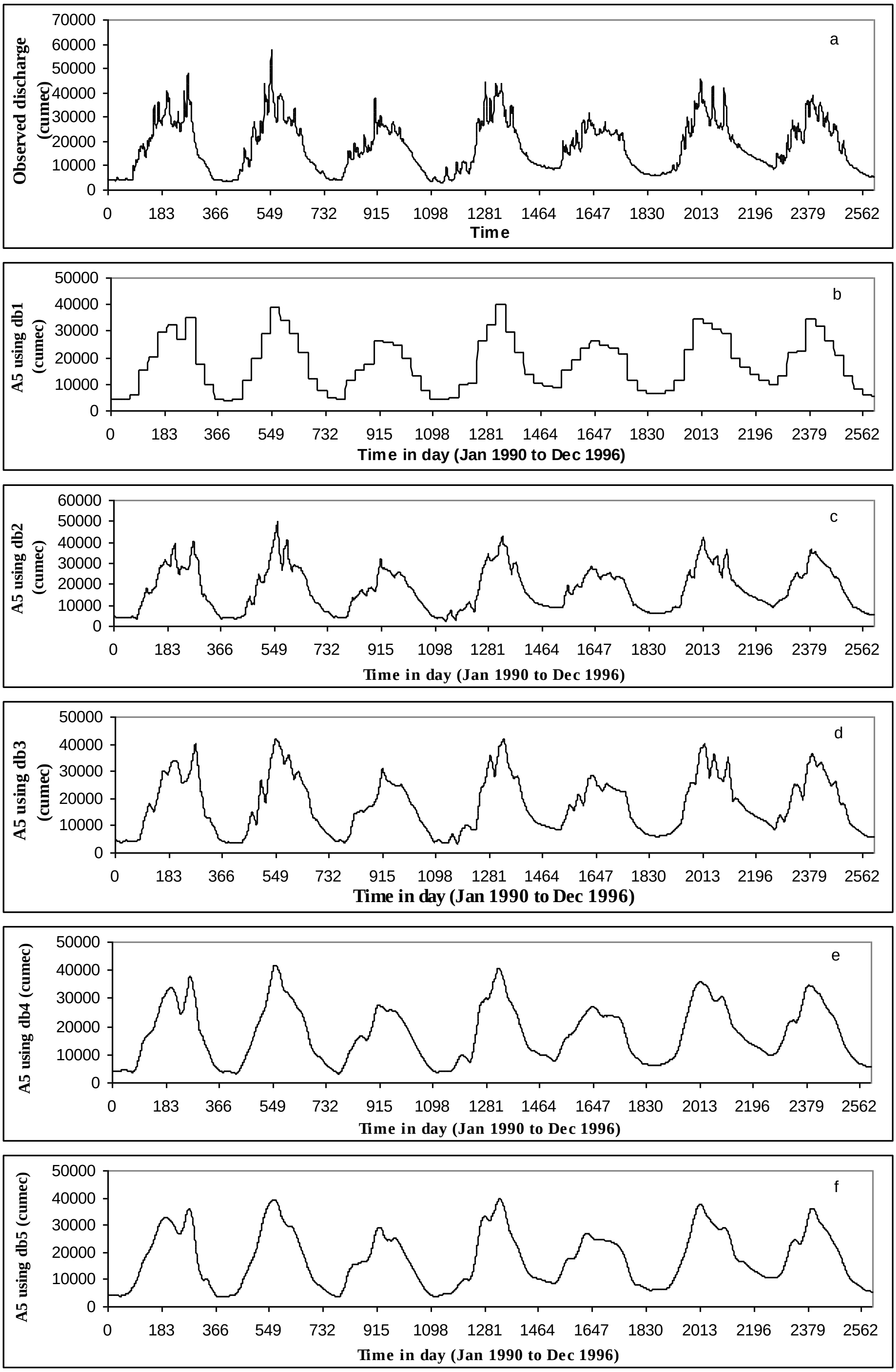


Fig. 4.1 (a) Observed daily flow (Pandu St.), Approximation coefficient at level 5 using (b) db1, (c) db2, (d) db3, (e) db4, (f) db5 mother wavelets

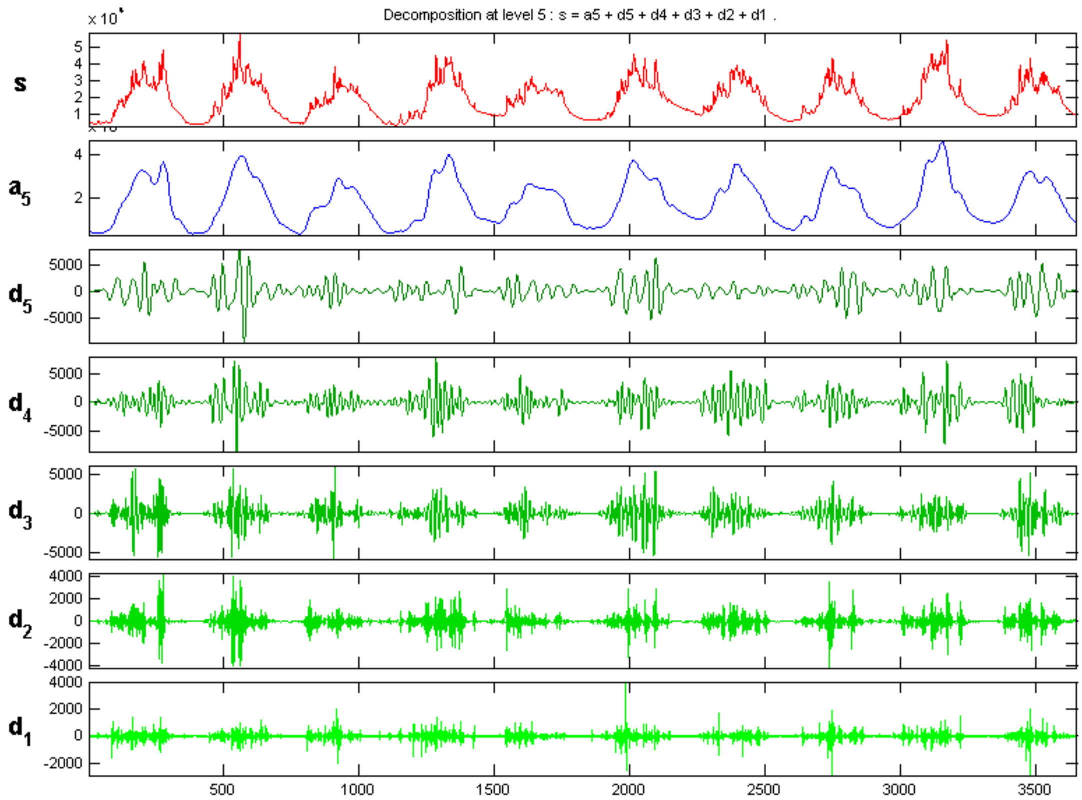


Fig. 4.2. Decomposition (approximation & detail) at level 5 using db5 wavelet

from 0.084 to 0.279. While BIAS values are found to very close to 1. It was found that for all WANN models, during training period BIAS values are slightly greater than 1 (indicating overestimation), while these are slightly less than 1 (indicating underestimation) during testing period of Pandu station. This may be due to extreme value lying in the training period.

For Pancharatna station (during training period), from the analysis, it was found that the value of R^2 decreased from 0.997 (db5*l*5 – WANN(db5)) for 2 day lead time to 0.968 (db5*l*6 – WANN(db5)) for 14 day lead time, while for ANN model this decrease was from 0.977 to 0.765. Also, the RMSE values increased from 652.40 (db5*l*5) to 2100.07 (db5*l*6) cumec for 2 day and 14 day lead time, respectively, while for ANN model this increase was from 1764.66 to 5704.97 cumec. The MAE values were found to increase from 413.09 (db5*l*5) to 1423.29 (db5*l*6) cumec for 2 day and 14 day lead time, respectively, while for ANN model this increase was from 1044.28 to 4035.67 cumec. The scatter index was found to increase from 0.040 (db5*l*5) to 0.129 (db5*l*6) for 2 day and 14 day lead time, respectively, while for ANN model this increase was from 0.109 to 0.353. Here, also, BIAS values are found to be very close to 1. For Pancharatna station (during testing period), from the analysis, it was found that the value of R^2 decreased from 0.994 (db5*l*5 – WANN(db5)) for 2 day lead time to 0.937 (db5*l*6 – WANN(db5)) for 14 day lead time, while for ANN model this decrease was from 0.960 to 0.673. Also, the RMSE values increased from 974.58 (db5*l*5) to 3100.53 (db5*l*6) cumec for 2 day and 14 day lead time, respectively, while for ANN model this increase was from 2463.33 to 7084.21 cumec. The MAE values were found to increase from 609.54 (db5*l*5) to 2027.85 (db5*l*6) cumec for 2 day and 14 day lead time, respectively, while for ANN model this increase was from 1401.83 to 5116.37 cumec. The scatter index was found to increase from 0.060 (db5*l*5) to 0.189 (db5*l*6) for 2 day and 14 day lead time, respectively, while for ANN model this increase was from 0.152 to 0.433. Here, also, BIAS values are found to be very close to 1. It was found that for all WANN models, during testing period BIAS values are slightly greater than 1 (indicating overestimation), while these are slightly less than 1 (indicating underestimation) during training period of Pancharatna station, which result is exactly opposite to Pandu station. This was due to extreme value lying in the testing period.

The above discussion and study of **Table 4.1 – 4.5** reveal that, in comparison with regular ANN model all WANN models have given better results for all lead times. The WANN model was found more accurate because wavelet transform decomposes the non-stationary time series data into several stationary approximation and details time series. In hybrid WANN model, wavelet transform takes care of non-stationarity while ANN handles non-linearity. Also river flow time series are characterized by non-linearity, non-stationarity and seasonality, which ANN models may not be able to cope without pre-processing of the input data.

4.2.3 Comparison among WANN Models With Different Daubechies Wavelets

In the present study, the results obtained by WANN models using db1 to db5 mother wavelets are also compared. Results of **Table 4.1 – 4.5** depict that WANN model with db5 mother wavelet [WANN(db5)] model has shown better performance compared to WANN(db1) to WANN(db4) models for both the stations. For 2 day lead time for Pandu station in training period (**Table 4.1 A**), from the analysis, it was found that the value of R^2 increased from 0.987 for db1*l*5 – WANN(db1) model to 0.995 for db5*l*5 – WANN(db5) model, while for Pancharatna station (**Table 4.1 B)** this increase was from 0.989 to 0.997 for the same models. Also, for Pandu station the RMSE values decreased from 1189.69 for db1*l*5 – WANN(db1) model to 741.35 cumec for db5*l*5 – WANN(db5), while for Pancharatna station this decrease was from 1199.43 to 652.40 cumec for the same models. For Pandu station the MAE values were found to decrease from 706.26 for db1*l*5 – WANN(db1) model to 473.51 cumec for db5*l*5 – WANN(db5), while for Pancharatna station this decrease was from 757.47 to 413.09 cumec for the same models. For Pandu station the scatter index was found to decrease from 0.067 for db1*l*5 – WANN(db1) model to 0.041 for db5*l*5 – WANN(db5), while for Pancharatna station this decrease was from 0.074 to 0.040. As discussed earlier, BIAS values are found to be very close to 1 for all the models. For 2 day lead time for Pandu station in testing period (**Table 4.1 A**), from the analysis, it was found that the value of R^2 increased from 0.987 for db1*l*5 – WANN(db1) model to 0.995 for db5*l*5 – WANN(db5) model, while for Pancharatna station (**Table 4.1 B)** this increase was from 0.979 to 0.994 for the same models. Also, for Pandu station the RMSE values decreased from 1220.89 for db1*l*5 – WANN(db1) model to 774.72 cumec for db5*l*5 – WANN(db5), while for Pancharatna station this decrease was from 1795.99 to 974.58 cumec for the same models. For Pandu station the MAE values were found to decrease from 733.79 for db1*l*5 – WANN(db1) model to 470.21 cumec for db5*l*5 – WANN(db5), while for Pancharatna station this decrease was from 1144.51 to 609.54 cumec for the same models. For Pandu station the scatter index was found to decrease from 0.063 for db1*l*5 – WANN(db1) model to 0.039 for db5*l*5 – WANN(db5), while for Pancharatna station this decrease was from 0.110 to 0.060. As discussed earlier, BIAS values are found to be very close to 1 for all the models.

The above discussion and the careful study of **Table 4.1 A** and **B,** reveal that WANN model's forecasting performance increases with increase in wavelet order, giving best results for db5 mother wavelet for both the stations and similar performance was observed for other lead times also. From the observed time series flow data (**Fig. 3.8 & 3.9, pp 38)**, it seems to have high frequency during monsoon season (June to September i.e. 4 months of every year), while comparatively low frequency in non-monsoon season (i.e. during remaining 8 months). The wavelets having wider support are capable of capturing low frequencies. On the other hand, wavelets having smaller support are capable of capturing high frequencies. db5 wavelet support width is 9 (see **Fig. 3.13, pp 48)** (support width of Daubechies wavelet is equal to 2 *i* - 1, where *i* = Daubechies wavelet order (Misiti, M. et al. 2010)), while db1, db2, db3 and db4 wavelets have support widths 1, 3, 5 and 7, respectively. In short, the db5 wavelet has a reasonable support and also has good time-frequency localization property and these together enable the model to capture both the underlying trend **(Fig. 4.1)** as well as the short term variablities in the time series better than db1 to db4 wavelet based forecast model. This observation is in congruence with the results obtained by Nourani et al. (2013), who showed that higher order mother wavelet (db4) provided comparatively better outcomes than lower order Haar (db1) wavelet. Another reason for having better results for db5 mother wavelet could be that its form is similar to the observed runoff signal fluctuation. **Figure 4.3** shows effect of Daubechies wavelet order on determination coefficient (R^2) for lead times 3, 7, and 14 day for Pandu station. **Fig. 4.3** reveals that, for the all the lead times R^2 is maximum for db5 mother wavelet.

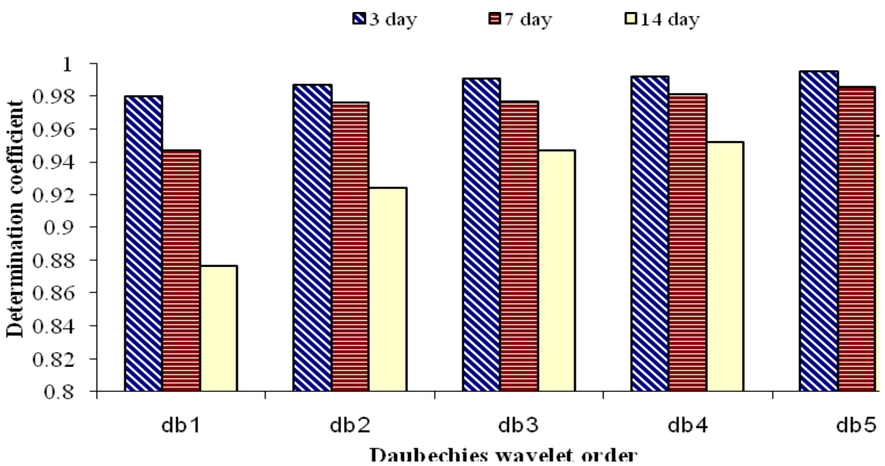


Fig. 4.3. Effect of Daubechies wavelet order on determination coefficient (R^2) for lead times 3, 7, and 14 day for Pandu station

To assess the potential of WANN model to preserve statistical properties, an analysis is carried out for testing period, the results of which are shown in the **Table 4.6** and **4.7** below. From these tables it was revealed that the statistical properties (mean, standard deviation and skewness coefficient) for almost all WANN models were same as that of the observed one except for some models in higher lead time (e.g. for WANN(db1) model for lead time 14 day, S_d = 11699 cumec, while the observed is 12388 cumec). Also, from the **Tables 4.6 and 4.7**, it was observed that the peak was fairly well estimated by WANN(db3) model (the error was limited 14.94 % for Pandu station and 23.31 % for Pancharatna station for highest lead time i.e. 14 day). However, the overall performance of WANN(db5) model is better. This may be due to spike like feature of db3 wavelet which is similar to sudden spikes of observed peak. As far as % error in low flow computation is concerned, almost all WANN models have worked fairly for Pandu station (error was limited to -14.22 %), but the error is high for Pancharatna station. This is due to peak lies in testing period for Pancharatna station. Time lag to peak was, also, fairly computed by all WANN models both the stations.

Table 4.6. Statistical properties of the observed and computed flow using WANN models during testing period for Pandu Station

Parameter	Lead time (day)	Observed	WANN (db1)	WANN (db2)	WANN (db3)	WANN (db4)	WANN (db5)
Q_{mean} (m^3/s)	2	19426	19356	19391	19445	19407	19395
	3		19383	19437	19397	19408	19435
	4		19443	19479	19439	19432	19458
	7		19544	19454	19453	19423	19436
	14		19454	19154	19645	19337	19564
S_d (m^3/s)	2	10662	10568	10573	10673	10735	10634
	3		10584	10619	10544	10593	10668
	4		10482	10665	10697	10609	10626
	7		10531	10501	10664	10718	10584
	14		10285	10098	10488	10156	10379
C_x	2	0.698	0.622	0.637	0.666	0.688	0.642
	3		0.644	0.654	0.635	0.626	0.714
	4		0.642	0.651	0.663	0.637	0.663
	7		0.647	0.667	0.704	0.734	0.718
	14		0.569	0.462	0.609	0.491	0.549
% Error in peak*	2	54100	11.55	8.06	4.44	5.10	8.81
	3		8.96	8.79	10.76	8.82	2.98
	4		6.56	10.6	4.81	11.46	8.19
	7		12.71	18.27	7.98	13.34	8.43
	14		14.35	27.49	14.94	19.58	18.62
% Error in low flow computation*	2	5567	1.31	3.70	2.75	2.13	3.16
	3		-1.60	5.01	2.62	3.89	1.34
	4		-4.69	1.53	-3.23	1.04	-3.82
	7		-3.74	6.05	3.59	3.00	5.32
	14		-14.22	-11.08	5.17	-9.55	-0.25
Time lag to peak (day)#	2	-	0	2 L	0	1A	1 L
	3		0	2 L	0	0	0
	4		0	0	0	0	0
	7		1 L	0	1 L	10 L	0
	14		1 A	15 A	21 A	7 A	1 L

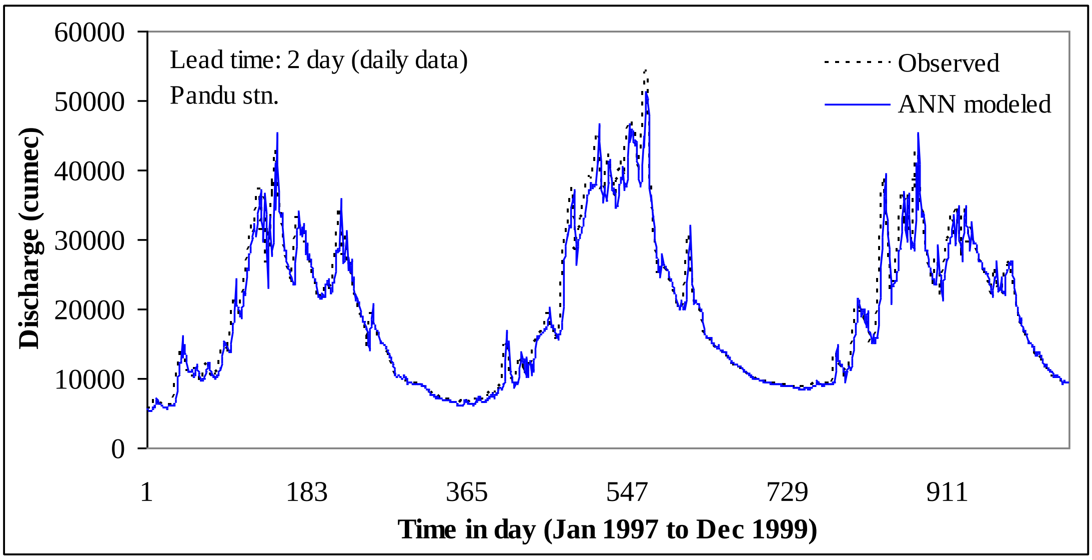
* positive – underestimation, negative – overestimation; # L =latter, A = ahead

Table 4.7. Statistical properties of the observed and computed flow using WANN models during testing period for Pancharatna Station

Parameter	Lead time (day)	Observed	WANN (db1)	WANN (db2)	WANN (db3)	WANN (db4)	WANN (db5)
Q_{mean} (m^3/s)	2	16236	16528	16359	16356	16343	16340
	3		16315	16331	16493	16320	16339
	4		16535	16381	16297	16438	16366
	7		16530	16663	16480	16402	16384
	14		16933	16960	16852	16372	16560
S_d (m^3/s)	2	12388	12140	12209	12192	12234	12187
	3		11972	12234	12266	12176	12150
	4		11814	12190	12187	12203	12184
	7		11553	12495	12110	12131	12152
	14		11699	12113	12353	12278	12051
C_x	2	0.968	0.931	0.961	0.954	0.960	0.973
	3		0.894	0.980	0.999	0.934	0.962
	4		0.854	0.955	0.939	0.929	0.944
	7		1.073	1.228	0.963	0.932	0.912
	14		0.986	0.947	0.964	0.996	0.842
% Error in peak*	2	76236	16.23	9.97	14.00	11.97	12.09
	3		23.97	9.79	8.88	20.69	11.10
	4		33.49	14.32	16.23	20.99	16.04
	7		33.70	1.36	23.90	23.11	19.09
	14		41.30	26.25	23.31	23.52	32.10
% Error in low flow computation*	2	1723	-161.34	-139.12	-138.30	-125.71	-139.64
	3		-138.59	-132.03	-150.66	-140.80	-135.63
	4		-177.77	-134.24	-132.33	-143.01	-136.16
	7		-237.72	-200.29	-175.39	-137.67	-140.51
	14		-241.67	-148.17	-208.71	-126.46	-135.93
Time lag to peak (day)#	2	-	2 L	1 L	1 L	1 L	1 L
	3		2 L	2 L	0	1 L	1 L
	4		2 L	3 L	1 L	1 L	2 L
	7		3 L	2 L	0	2 L	1 L
	14		8 L	5 L	1 A	2 L	4 L

* Positive – underestimation, negative – overestimation; # L =latter, A = ahead

The time series and scatter plots for ANN and WANN(db5) models for all lead times (i.e. 2, 3, 4, 7, and 14 day) for daily time series data are shown in the **Fig. 4.4 to Fig. 4.13**.



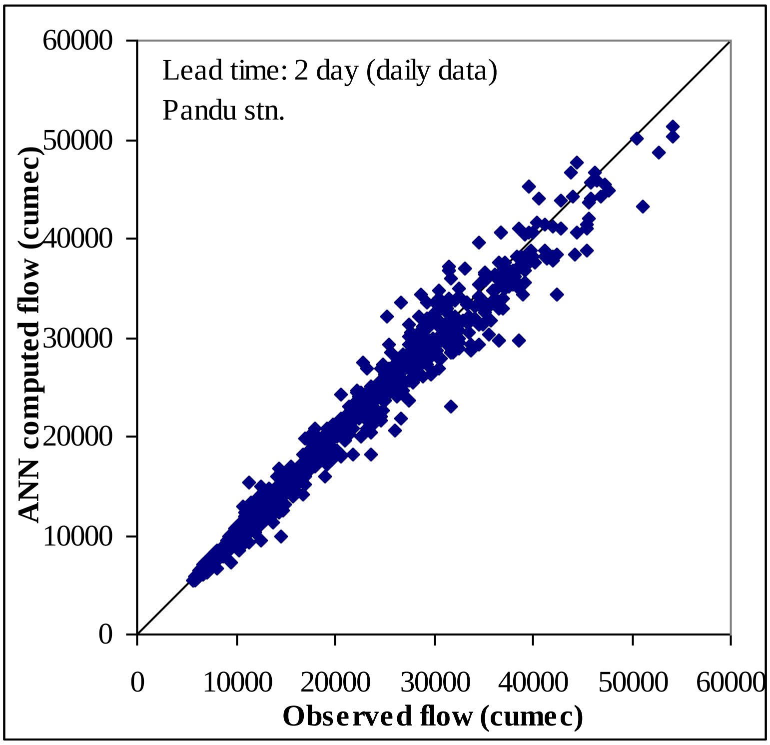
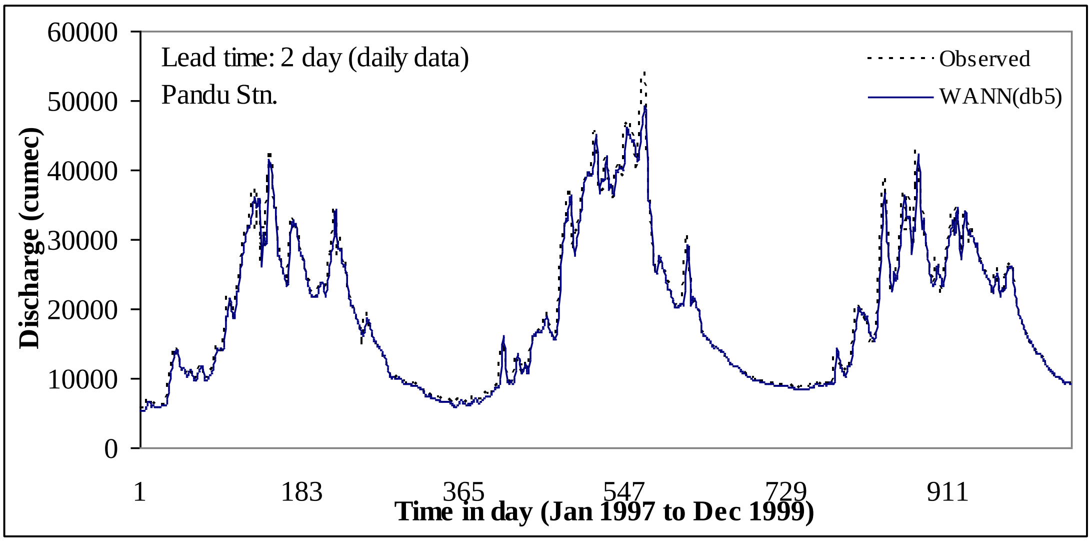


Fig.4.4. Flow series and scatter plot between observed and ANN modeled flow for 2 day lead time (daily data) for Pandu station during testing period



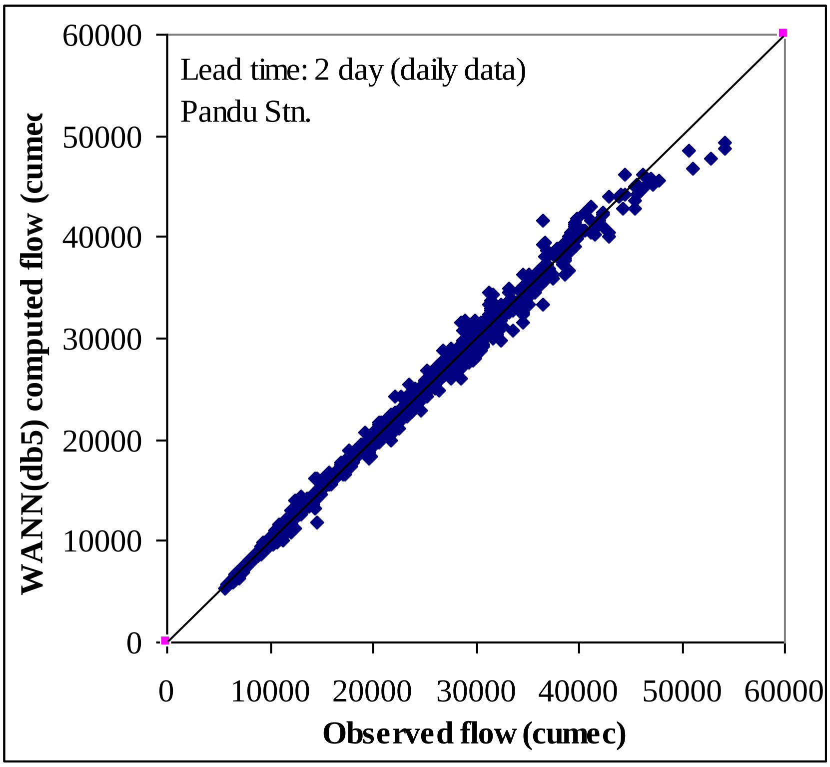
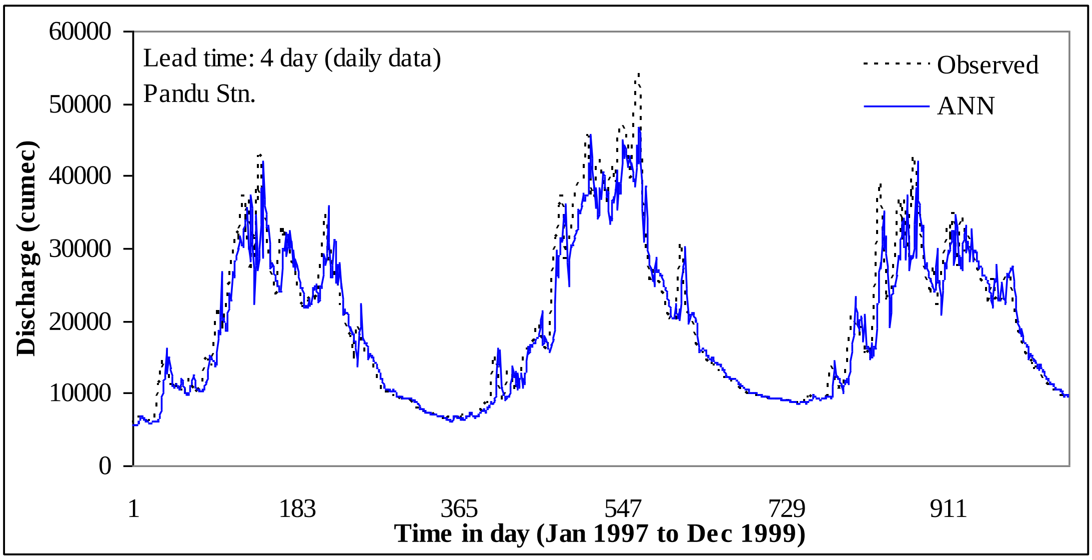


Fig.4.5. Flow series and scatter plot between observed and WANN(db5) modeled flow for 2 day lead time (daily data) for Pandu station during testing period



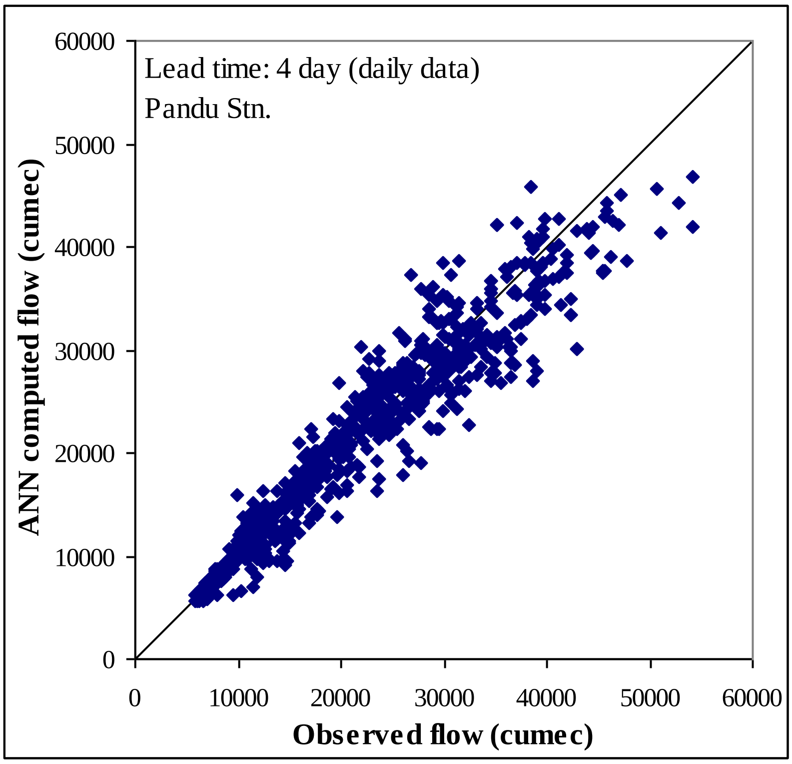
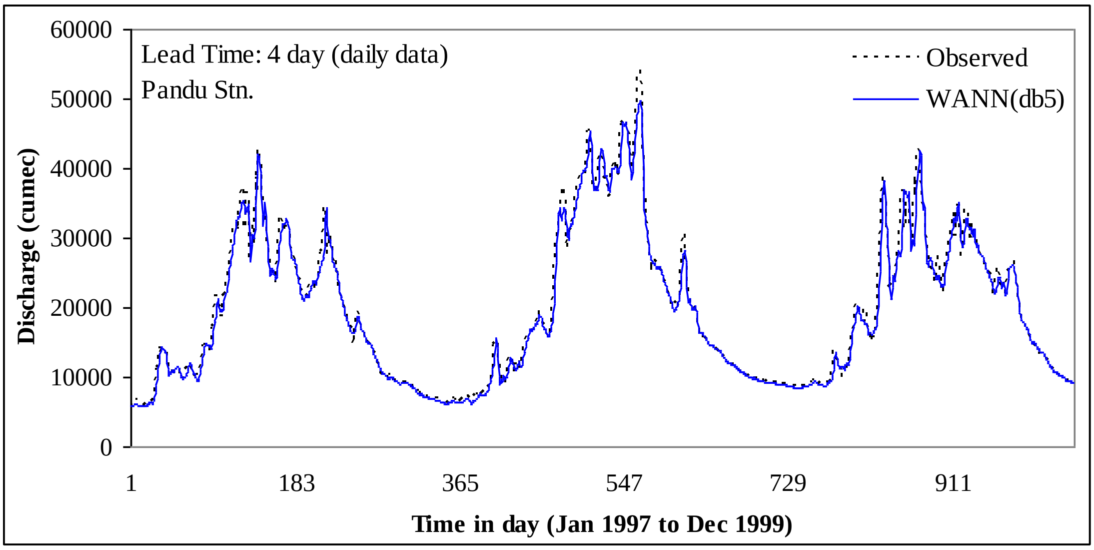


Fig.4.6. Flow series and scatter plot between observed and ANN modeled flow for 4 day lead time (daily data) for Pandu station during testing period



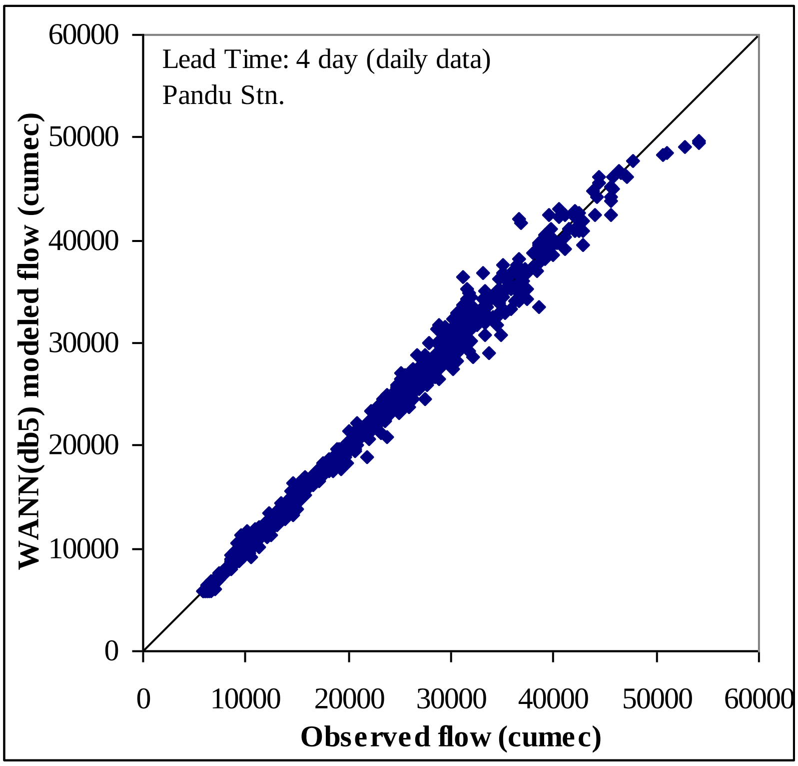
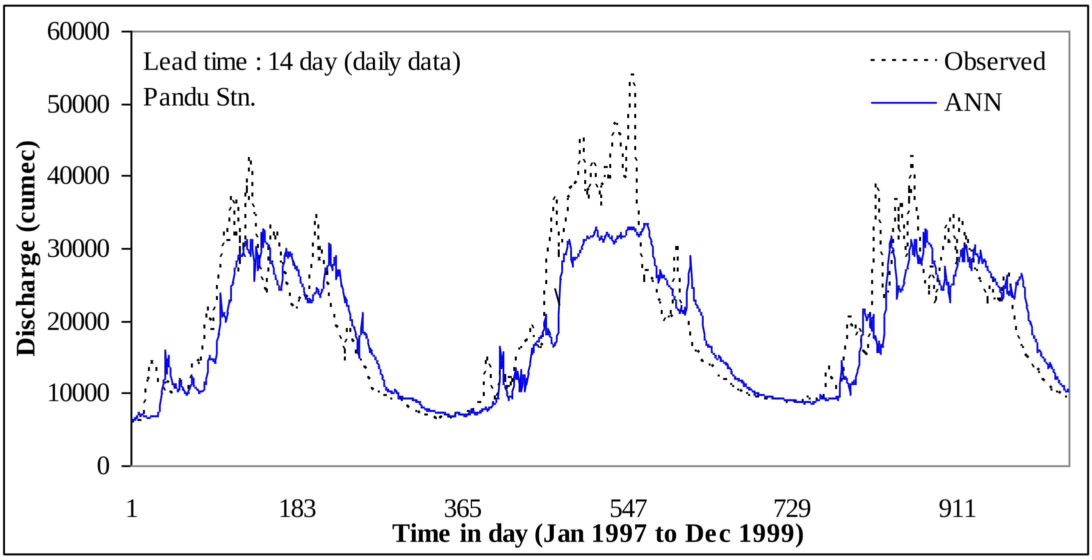


Fig.4.7. Flow series and scatter plot between observed and WANN(db5) modeled flow for 4 day lead time (daily data) for Pandu station during testing period



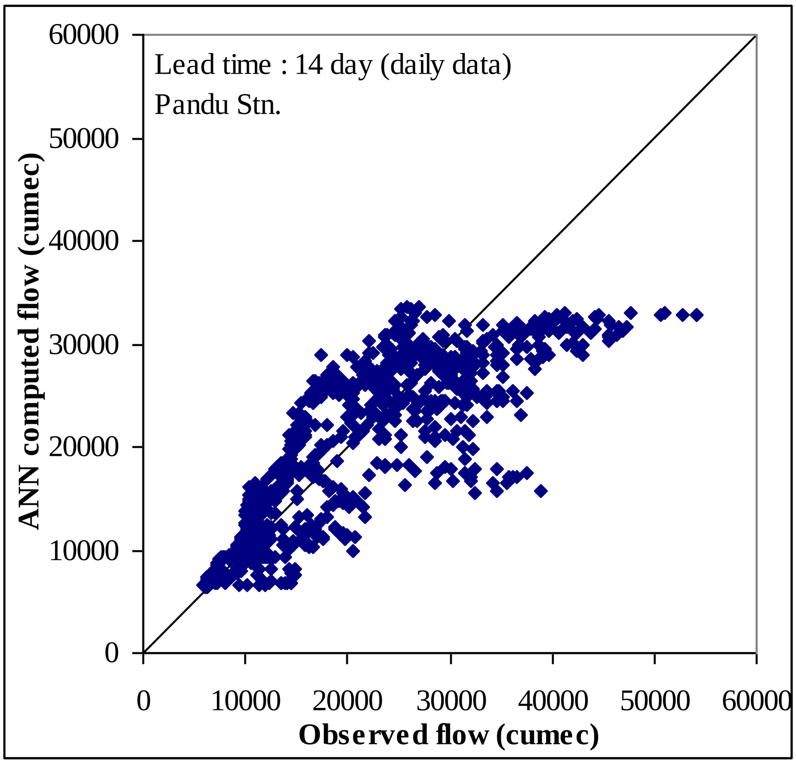
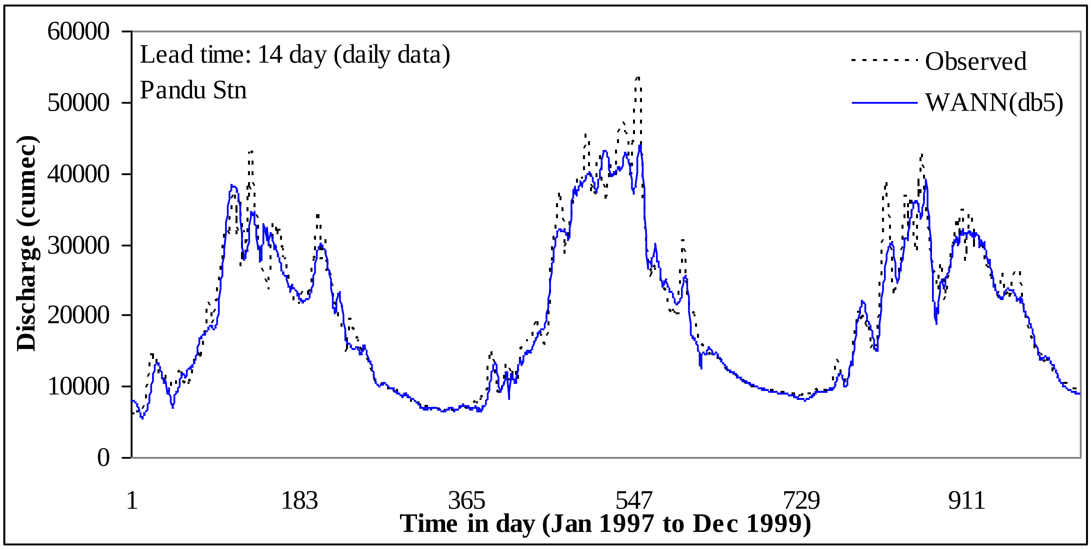


Fig.4.8. Flow series and scatter plot between observed and ANN modeled flow for 14 day lead time (daily data) for Pandu station during testing period



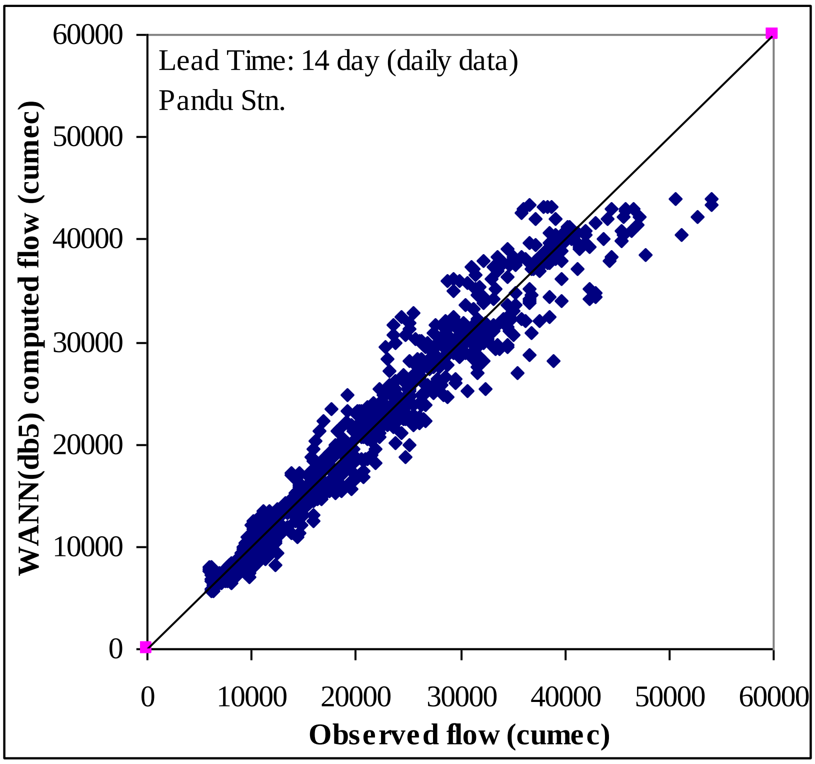
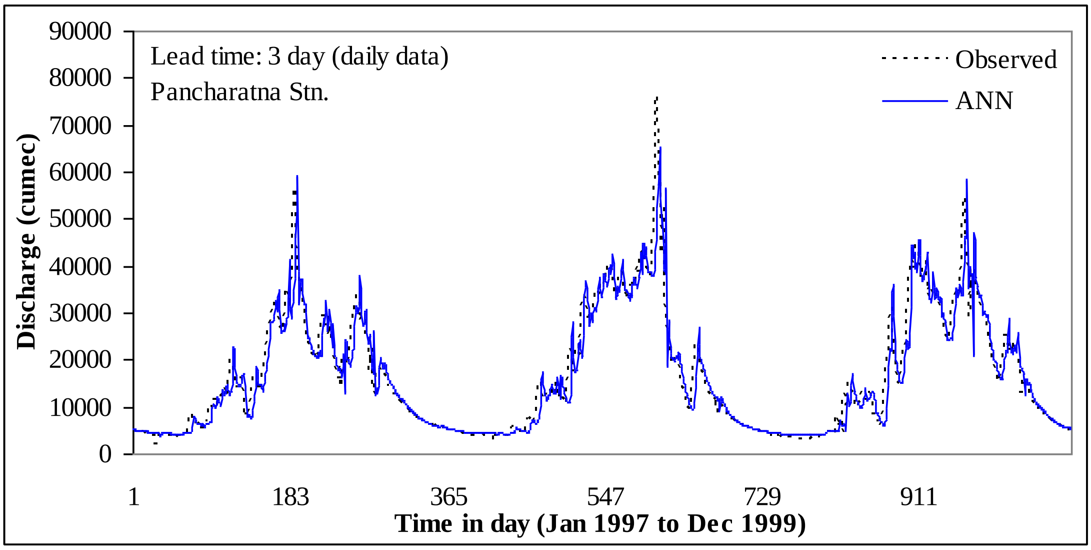


Fig.4.9. Flow series and scatter plot between observed and WANN(db5) modeled flow for 14 day lead time (daily data) for Pandu station during testing period



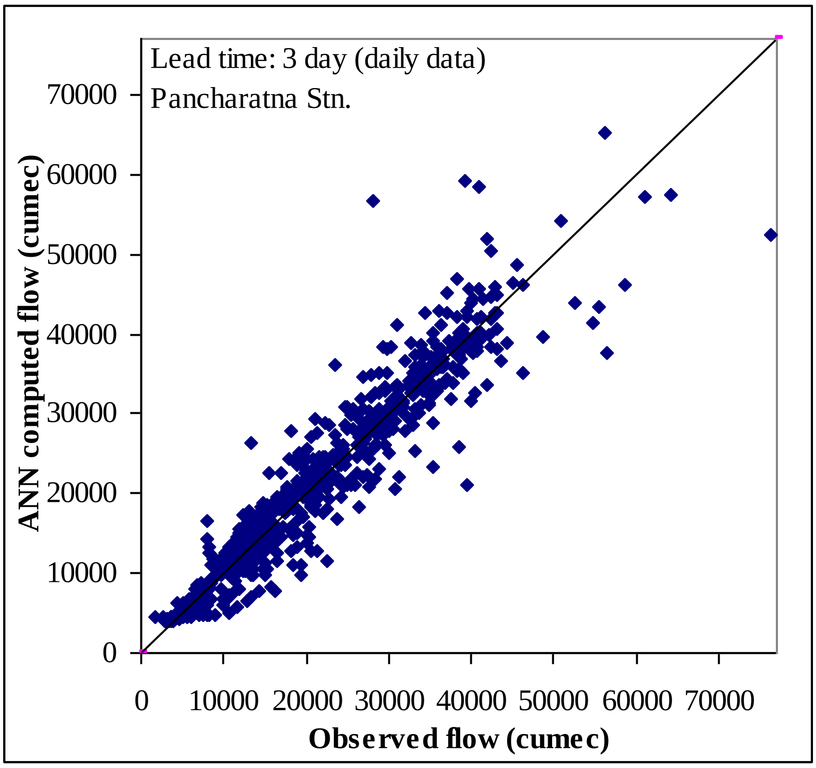
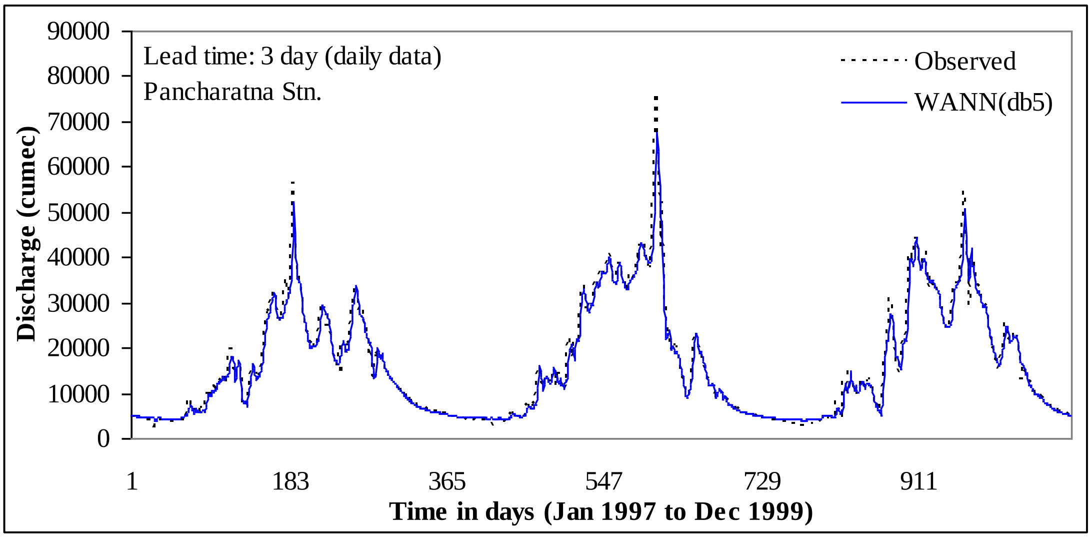


Fig.4.10. Flow series and scatter plot between observed and ANN modeled flow for 3 day lead time (daily data) for Pancharatna station during testing period



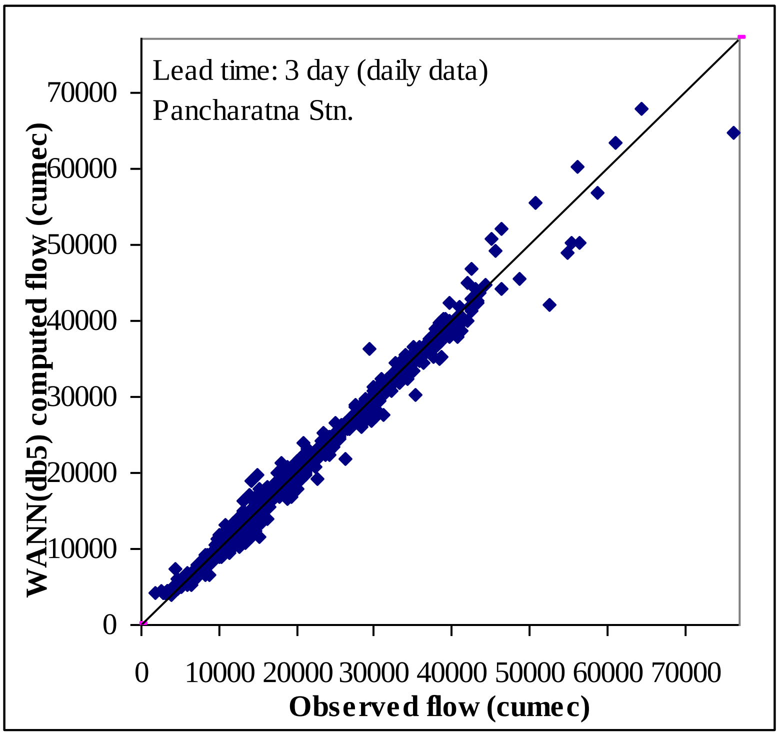
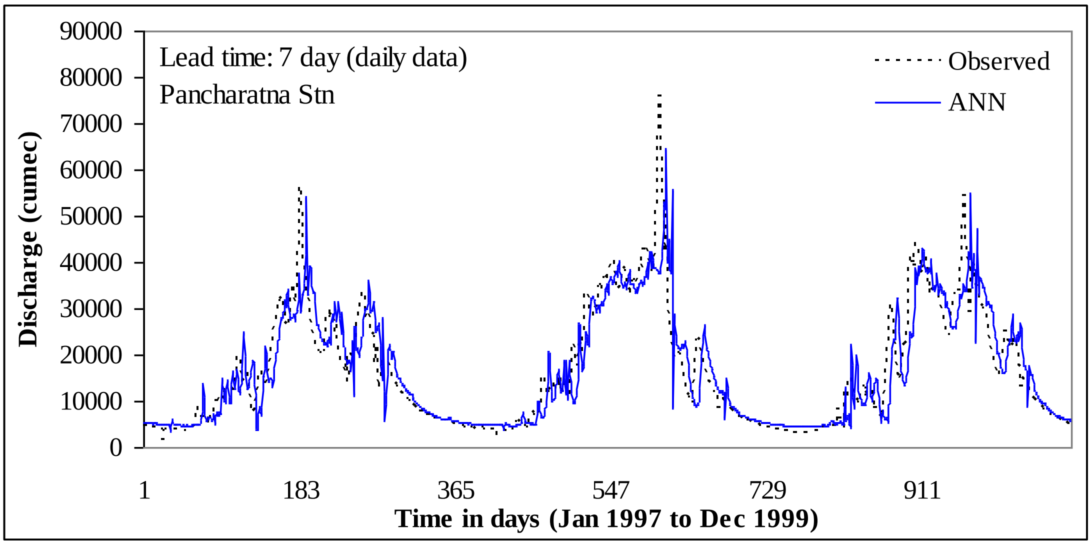


Fig.4.11. Flow series and scatter plot between observed and WANN(db5) modeled flow for 3 day lead time (daily data) for Pancharatna station during testing period



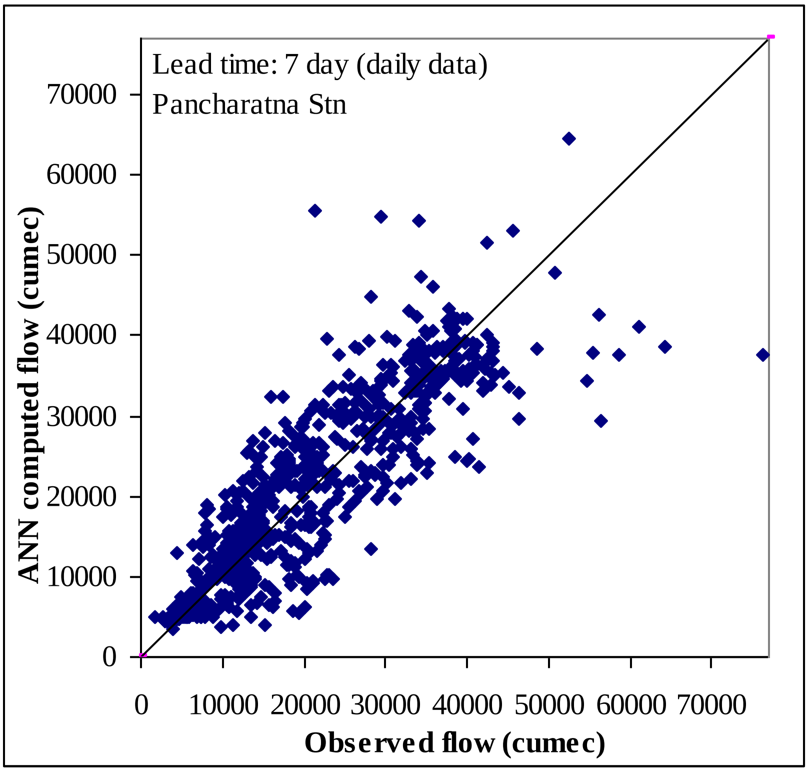
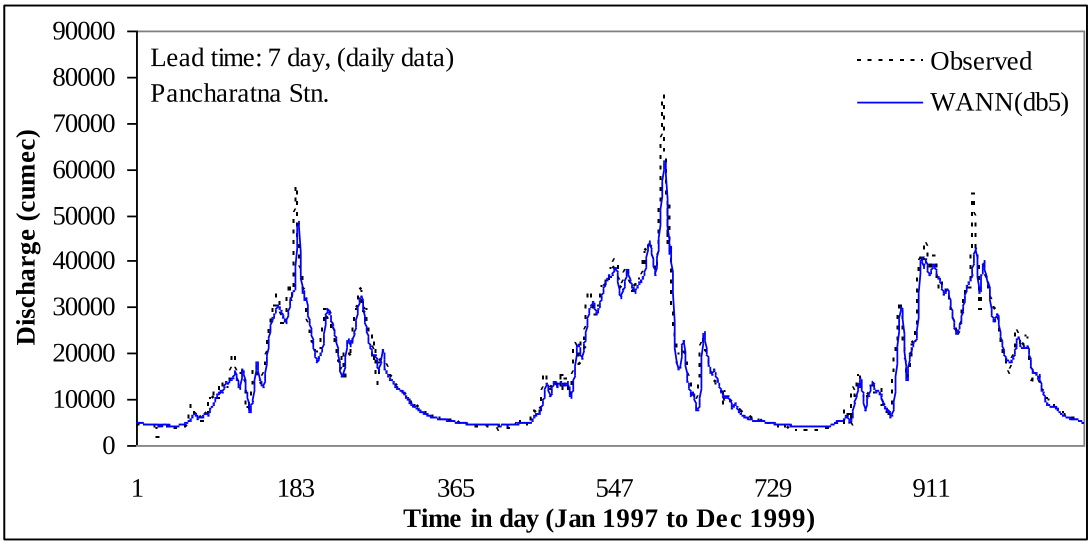


Fig.4.12. Flow series and scatter plot between observed and ANN modeled flow for 7 day lead time (daily data) for Pancharatna station during testing period



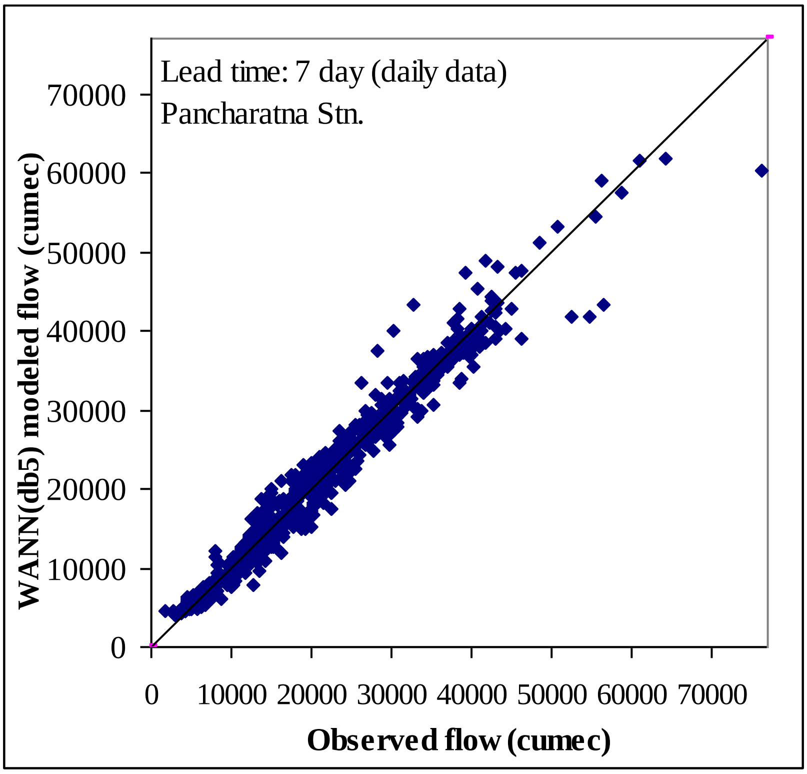


Fig.4.13. Flow series and scatter plot between observed and WANN(db5) modeled flow for 7 day lead time (daily data) for Pancharatna station during testing period

Figure 4.4 to Figure 4.13 shows flow series and scatter plots for ANN and WANN(db5) models for various lead times for Pandu and Pancharatna stations. In these figures the results of best WANN model (i.e. WANN(db5)) were compared with ANN models. From these figures it was observed that the performance of WANN(db5), for all lead times and for both the stations, was superior compared to ANN models.

From **Figure 4.5** (pp 68) (lead time: 2 day, Pandu stn, WANN(db5)), it was observed (from flow series plot) that WANN(db5) model was able to capture multiple peaks during all the three monsoon seasons. However, it was not able to capture highest peak during the monsoon season of June – September, 1998. The small multiple peaks in the rising limbs (occurred due to melting of snow in the Himalaya), as well as in the falling limbs, were well captured by the model. **Figure 4.5** also depicts that, the model was able to capture low flows precisely. From the scatter plot of **Fig. 4.5**, it was observed that most of the points were very close to 45^{o} line (SI = 0.039), with few points having higher magnitudes of observed flow on lower side, indicating models underestimation (BIAS = 0.998). The models performance was very satisfactory (R^2 = 0.995, RMSE = 774.72 cumec, MAE = 470.21 cumec) (Dawson and Wilby, 2001). Similar trend was observed in **Fig. 4.7** (pp 70) (lead time: 4 day, Pandu stn., WANN(db5)).

As shown in **Figure 4.9** (pp 72) (lead time: 14 day, Pandu stn, WANN(db5)), it was observed that, the multiple peaks in the monsoon season were not well captured by the model. Small multiple peaks in the rising and falling limb of hydrograph were, also, not captured properly. However, the model was able to capture low flows precisely (with % error = - 0.25). The overall models performance was very satisfactory (R^2 = 0.956, RMSE = 2228.35 cumec).

In this study it was found that, with increase in lead time models performance decreases. This may be due to weak link existing between input and output data. However, the hybrid WANN models have shown that the use of WT as a preprocessing technique can produce predictions more accurate and this is due to WT extracts periods and seasonality.

Figure 4.11 (pp 74) shows time series and scatter plots for lead time 3 day (Pancharatna station) for WANN(db5) model. Flow series plot reveals that, the model was able to capture multiple peaks in all monsoon seasons except the peak values. But it was not able to capture low flows more precisely. From the scatter plot, it was observed that the model computed values were close to observed with SI = 0.07, and BIAS = 1.006 (slightly overestimation). The models overall performance was very satisfactory (R^2 = 0.991, RMSE = 1144.08 cumec, MAE = 720.30 cumec)

Figure 4.13 (pp 76) shows flow series and scatter plots of WANN(db5) model for lead time 7 day (Pancharatna station). From time series plot, it was observed that, due to increase in lead time, it was not able to capture multiple peaks, however, the modeled and observed values are well matched in the rising and falling limbs of hydrograph. Also model was not able to capture low flows precisely. However, models overall performance was very satisfactory (R^2 = 0.979, RMSE = 1771.43 cumec, MAE = 1127.24 cumec, BIAS = 1.007, and, SI = 0.109).

As mentioned earlier, the results of Pancharatna station are found to be comparatively poor than Pandu station, because of high non-stationarity and peak of overall time series lies in testing period.

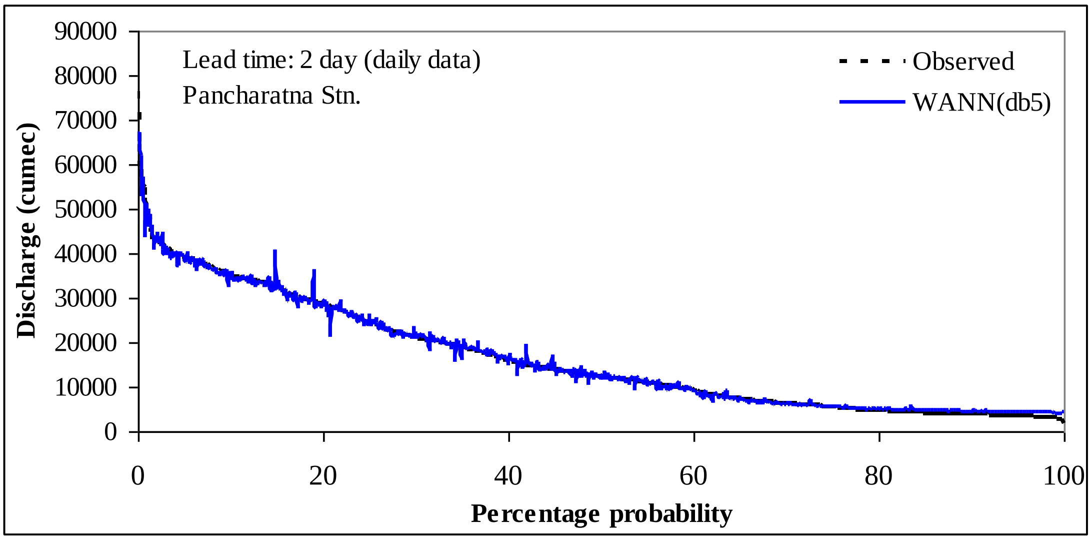
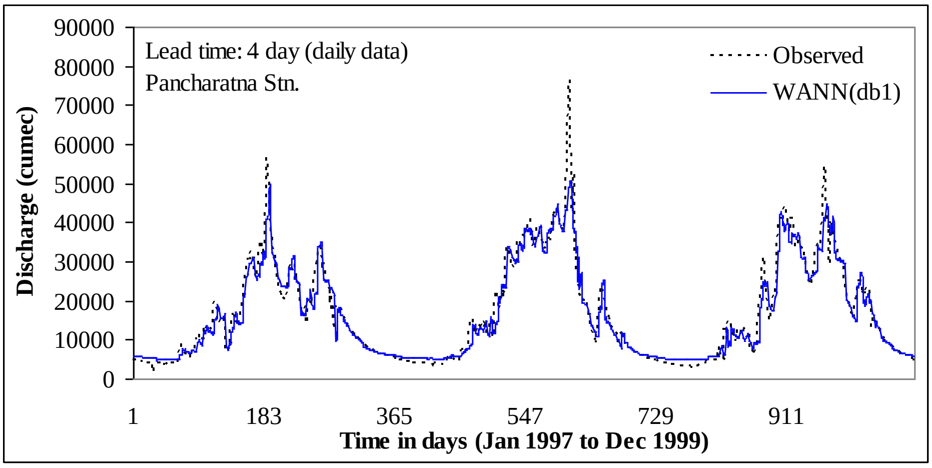


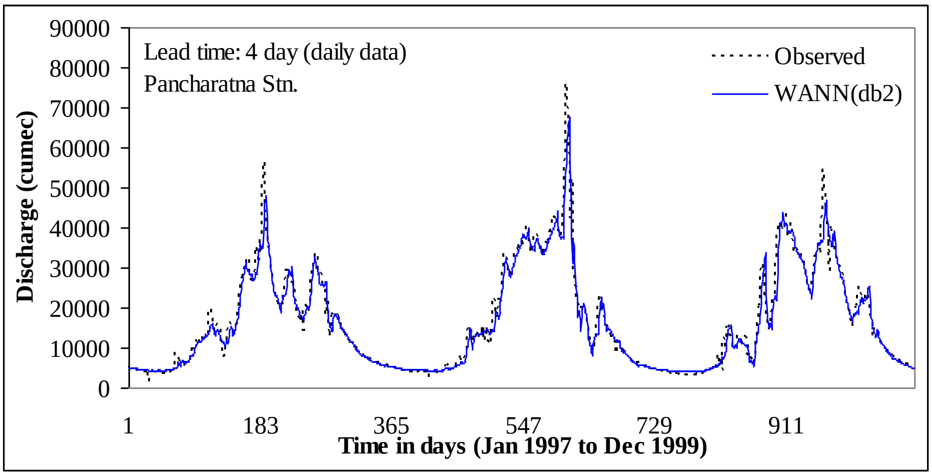
Fig. 4.14. Discharge-frequency curve during testing period for lead time 2 day (daily data), Pancharatna Stn.

Figure 4.14 shows discharge-frequency curve (also known as flow-duration curve) in testing period for 2 day lead time for Pancharatna station. The flow-duration curve represents the cumulative frequency distribution and can be considered to represent the streamflow variation of an average year. The discharge ordinate Q_t at any percentage probability P_t represents the flow magnitude in an average year that can be expected to be equalled or exceeded P_t percent of time and is termed P_t % dependable flow. This is very important in the planning of water resources projects, in flood control studies, etc. it may be noted from **Figure 4.14** that, WANN(db5) model captured high and low flows, but it was unable to capture medium flow (20,000 – 40,000 cumec).

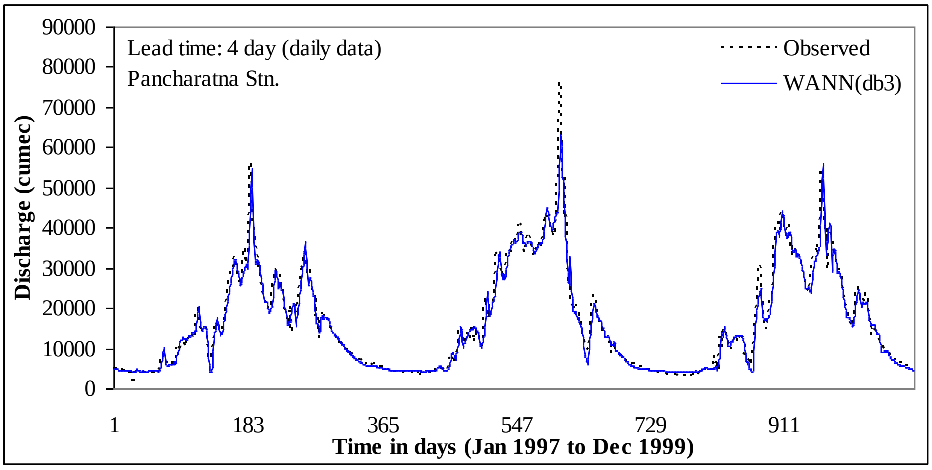
Figures 4.15 and 4.16 show flow series and scatter plots, respectively, for best WANN model (with low RMSE) for Daubechies wavelets of order db1 to db5 for lead time 4 day (Pancharatna station). From these plots it was observed that, with increase in wavelet order, the model's forecasting performance was increased. Also, it is worth to mention that, db3 wavelet (**Fig. 4.15 c**) was able to capture all the peaks in the monsoon season compared to other wavelets.



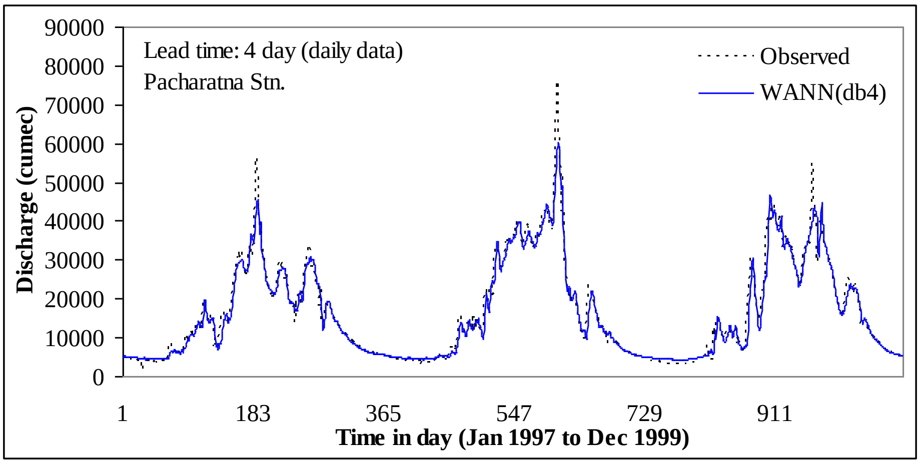
(a)



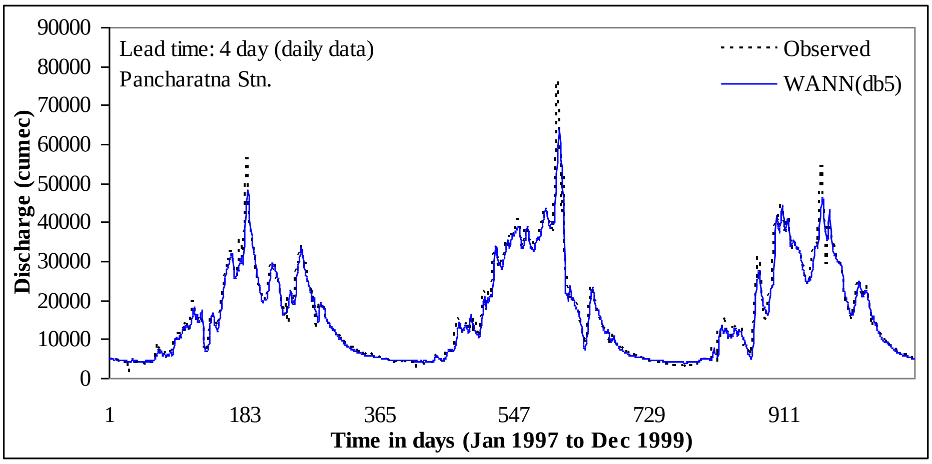
(b)



(c)



(d)



(e)

Fig.4.15. Flow series plots between observed and (a) WANN(db1), (b) WANN(db2), (c) WANN(db3), (d) WANN(db4), (e) WANN(db5) modeled flow for 4 day lead time (daily data) for Pancharatna station during testing period

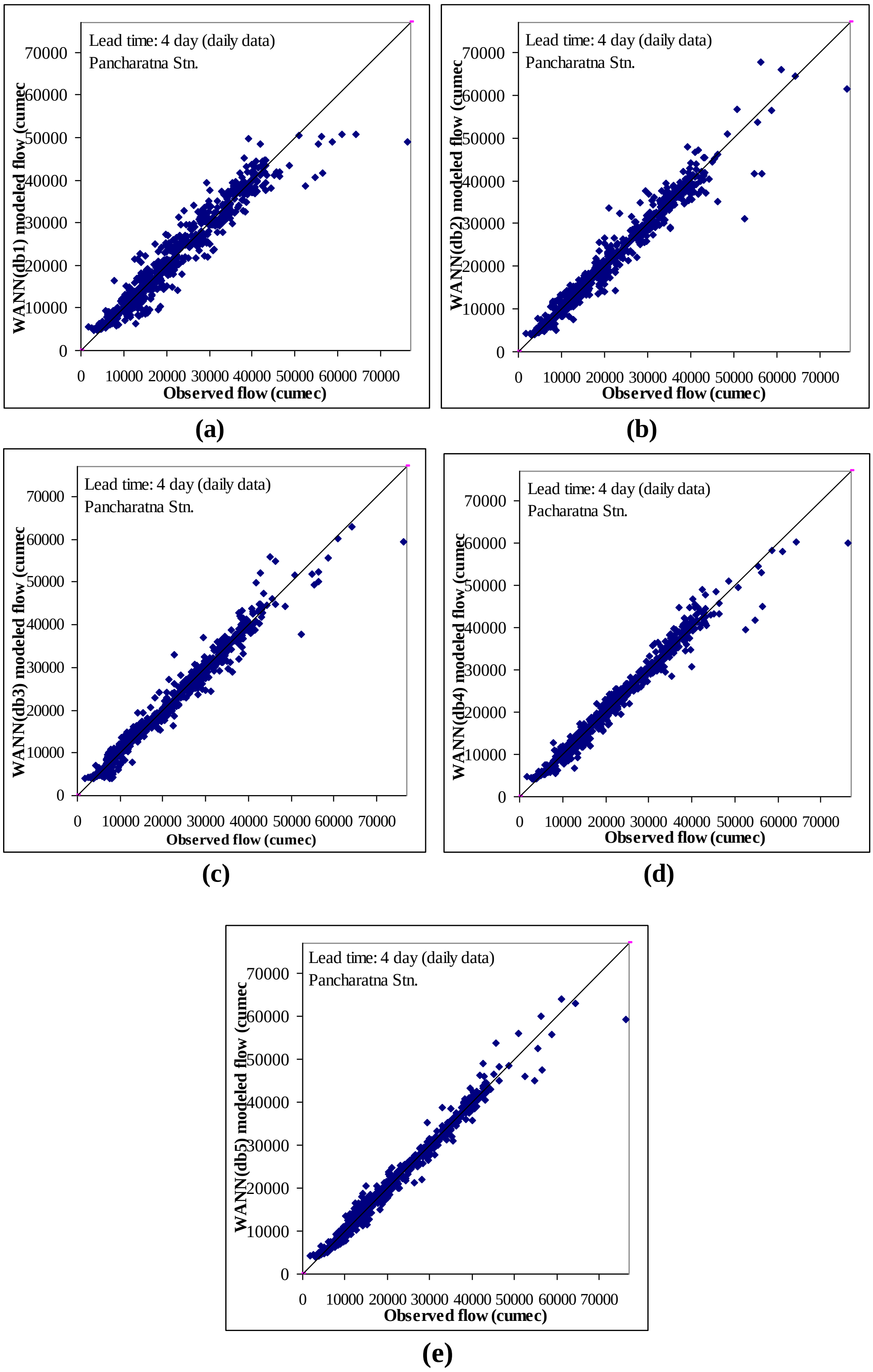


Fig.4.16. Scatter plots between observed and (a) WANN(db1), (b) WANN(db2), (c) WANN(db3), (d) WANN(db4), (e) WANN(db5) modeled flow for 4 day lead time (daily data) for Pancharatna station during testing period

4.2.4 Influence of Decomposition Level on Model Performance

It was also one of the main objectives of the present study, to assess the influence of decomposition level on WANN model performance. Deciding the optimal decomposition level of the time series data in wavelet analysis plays an important role in preserving the information and reducing the distortion of the datasets. However, there is no existing theory to specify the number of decomposition levels needed for any time series (Pandhiani et al. 2013). In the present study, the optimal level of decomposition was found by trial-and-error method as carried out by previous researchers (Deka and Prahlada, 2012, Moosavi et al. 2013).

As mentioned earlier, for daily time series data of length 2557 the maximum required number of decomposition levels are 11. However, in the present study, the normalized time series data is decomposed into maximum 7 levels. Because it was found that, the optimal level of decomposition (low RMSE) was found to lie between 4 and 6 (depending on lead time) for daily data.

Careful study of **Table 4.1 A – 4.5 A** (pp 53 to 57) reveals that, with increase in decomposition level model's efficiency increases upto a certain level, thereafter it's starts declining. As mentioned earlier, high frequency components of the original time series are captured in the first resolution level and with increase in decomposition level (scale) signal becomes more and more smoother (stationary) (see **Fig. 4.2,** pp 61**)**, hence prediction errors were not increased with scale. The other reason might be, at the optimum scale the stretched baby wavelet coinciding with the original signal. After reaching the optimal level, the model's forecasting performance decreased because with increase in decomposition level, the number of neurons in input layer increased, creating complex non linear relationship, which ANN may not be able to cope. From **Table 4.1 A – 4.5 A** it was observed that, for low lead time (2, 3, and 4 day) the optimum level was 4^{th} (2^4-day mode i.e. 2 week) and 5^{th} (2^5-day mode i.e. 1 month), while for higher lead time (7 and 14 day) it was 5^{th} (2^5 – day mode i.e. 1 month) and 6^{th} (2^6- day mode i.e. 2 month). It is worth to mention that, optimum level of decomposition increased with lead time. **Figure 4.17** shows the effect of decomposition level on RMSE for WANN(db4) and WANN(db5) models for lead time 4 day (Pandu station). For both the models the optimum decomposition level was found to be 5.

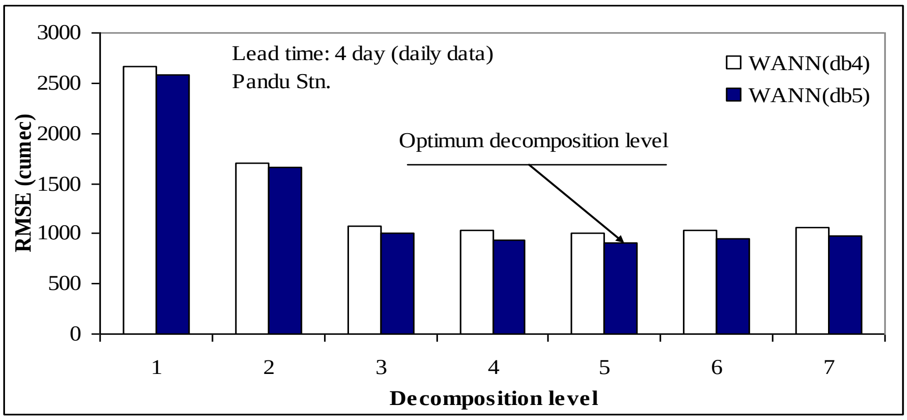


Fig.4.17. Effect of decomposition level on RMSE (lead time: 4 day, Pandu Stn.)

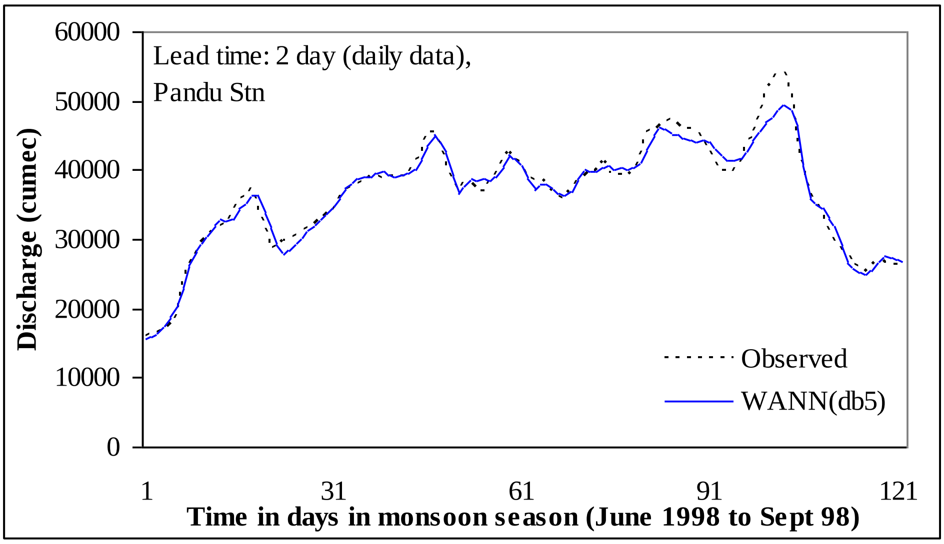
4.2.5 Analysis of Results for Monsoon Season of Testing Period

Finally, as the Brahmaputra River carries heavy flood in the monsoon season (June to September), attempt was made to investigate the model's accuracy during monsoon period (for three years in testing period 1997 to 1999) for WANN(db5) model. **Table 4.8** presents the values of statistical parameters for WANN(db5) model for monsoon season of testing period. **Table 4.8** depicts that, WANN(db5) model performed well for monsoon season giving very satisfactory results (except in case the of highest lead time 14 day), despite of high non-stationarity.

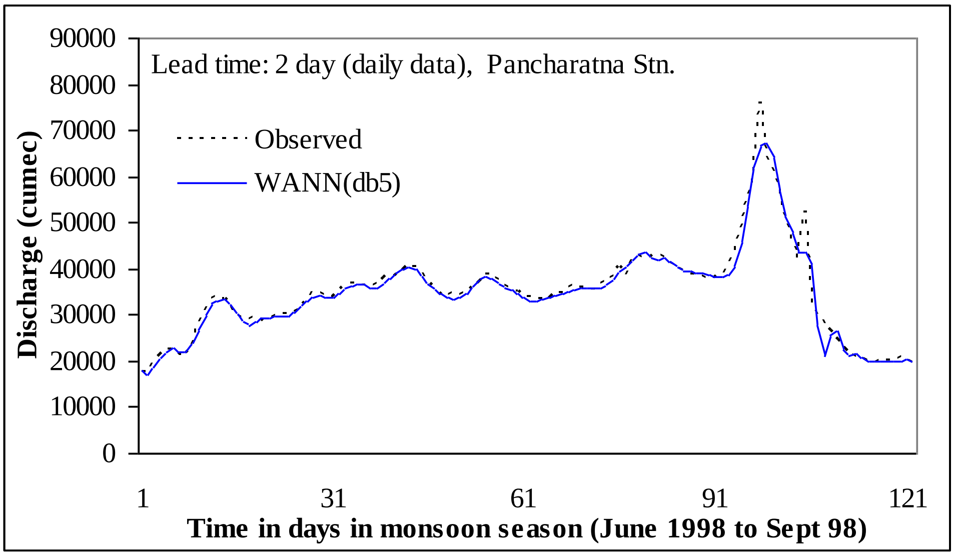
Table 4.8. Values of statistical parameters for WANN(db5) model for monsoon season (June to September) in testing period

Year	Pandu Station					Pancharatna Station				
	RMSE	R^2	MAE	BIAS	S.I.	RMSE	R^2	MAE	BIAS	S.I.
Lead time: 2 day										
1997	1002.92	0.962	637.45	1.000	0.035	1019.43	0.980	684.94	1.000	0.039
1998	1359.43	0.970	964.25	0.998	0.037	1756.88	0.967	892.63	0.994	0.050
1999	1150.23	0.936	864.99	0.999	0.039	1231.57	0.976	768.97	0.998	0.039
Lead time: 3 day										
1997	1280.02	0.938	814.88	1.000	0.045	1302.26	0.968	789.61	1.000	0.049
1998	1198.65	0.977	907.27	1.002	0.033	1949.73	0.960	1077.25	0.992	0.056
1999	971.65	0.954	735.87	0.997	0.033	1483.78	0.965	966.63	0.996	0.047
Lead time: 4 day										
1997	1344.17	0.931	997.9	1.000	0.047	1457.32	0.960	948.69	1.000	0.056
1998	1361.19	0.970	1016.36	0.996	0.037	2460.41	0.936	1374.27	0.995	0.071
1999	1531.10	0.887	1075.25	1.001	0.052	1711.08	0.953	1148.83	0.999	0.054
Lead time: 7 day										
1997	1570.51	0.906	1192.17	1.000	0.056	2413.60	0.889	1679.33	1.000	0.092
1998	1947.61	0.939	1507.31	0.986	0.053	2992.62	0.905	1855.51	1.000	0.086
1999	2268.05	0.752	1636.47	0.997	0.077	2241.25	0.920	1577.54	1.000	0.071
Lead time: 14 day										
1997	3130.39	0.628	2334.03	1.000	0.111	3998.42	0.697	3097.05	1.000	0.153
1998	3605.32	0.791	2624.24	0.991	0.098	4771.57	0.759	3122.96	1.012	0.137
1999	3264.38	0.486	2458.46	0.998	0.111	4954.74	0.608	4087.51	0.996	0.157

Note: RMSE and MAE are in cumec unit.



(a)



(b)

Fig. 4.18. Flow series comparison between observed and WANN(db5) modeled flow for monsoon season (June 1998 to Sept 98) for lead time 2 day for (a) Pandu Stn. (b) Pancharatna Stn.

Figures 4.18 (a) and (b) show flow series plots between observed and WANN(db5) modeled for monsoon period for lead time 2 day for Pandu and Pancharatna stations, respectively. It was seen from these figures that, the model was able to capture all multiple peaks (except highest) fairly well.

4.3 MODEL RESULTS FOR WEEKLY TIME SERIES DATA

Optimum results (with low RMSE) of model runs carried out by varying the order of Daubechies wavelet and decomposition level for weekly time series data for both the stations are shown in following tables (**Table 4.9 to 4.12**). Results of ANN models are also shown in the same tables for comparison.

Table 4.9. Values of statistical parameters for ANN and WANN models Lead time: 1 week (Pandu Station - Weekly data))

Model type	Training period RMSE	R^2	MAE	BIAS	S.I.	Testing period RMSE	R^2	MAE	BIAS	S.I.	Optimum ANN structure
ANN	4253.35	0.835	2802.36	0.974	0.236	4971.35	0.791	3229.57	0.936	0.258	4-9-1
db1*l*1	3371.47	0.896	2335.92	1.028	0.187	3795.24	0.878	2349.46	1.009	0.197	8-3-1
db1*l*2	3073.10	0.914	2054.31	1.019	0.170	3313.25	0.907	2169.73	1.003	0.172	12-3-1
db1*l*3	2955.61	0.920	2014.22	1.005	0.164	3209.93	0.913	2047.63	0.996	0.166	16-3-1
db1*l*4	3032.01	0.916	2136.42	1.002	0.168	**3191.31**	**0.914**	**2204.58**	**0.986**	**0.165**	**20-3-1**
db1*l*5	3145.45	0.910	2195.76	1.012	0.174	3524.75	0.895	2380.46	0.994	0.183	24-3-1
db1*l*6	3100.37	0.912	2223.36	1.009	0.172	3615.08	0.889	2506.33	0.975	0.187	28-3-1
db1*l*7	3171.45	0.908	2142.37	1.006	0.176	6472.64	0.646	5456.67	0.730	0.336	32-3-1
db2*l*1	3006.36	0.918	2172.31	1.027	0.167	3177.57	0.915	2064.41	0.987	0.165	8-2-1
db2*l*2	2078.57	0.961	1524.58	1.011	0.115	2934.43	0.927	1747.16	1.009	0.152	12-2-1
db2*l*3	2091.44	0.960	1462.89	1.012	0.116	2727.80	0.937	1682.32	0.994	0.141	16-2-1
db2*l*4	2306.50	0.951	1653.72	0.997	0.128	**2665.58**	**0.940**	**1760.14**	**0.978**	**0.138**	**20-2-1**
db2*l*5	2336.56	0.950	1649.48	1.004	0.129	2784.72	0.934	1925.27	0.978	0.144	24-2-1
db2*l*6	2398.07	0.947	1668.14	1.031	0.133	2798.72	0.934	1846.12	1.023	0.145	28-2-1
db2*l*7	2568.22	0.939	1967.25	1.037	0.142	3041.83	0.922	2278.74	1.035	0.158	32-2-1
db3*l*1	2309.94	0.951	1675.72	1.013	0.128	2475.77	0.948	1692.69	0.988	0.128	8-2-1
db3*l*2	1814.14	0.970	1310.14	0.997	0.100	1873.59	0.970	1323.13	0.985	0.097	12-2-1
db3*l*3	1778.55	0.971	1270.50	1.000	0.099	1787.55	0.973	1317.73	0.984	0.093	16-2-1
db3*l*4	1757.86	0.972	1152.84	1.004	0.097	**1682.11**	**0.976**	**1198.33**	**0.998**	**0.087**	**20-2-1**
db3*l*5	1861.91	0.968	1325.83	1.005	0.103	1854.61	0.971	1338.71	0.998	0.096	24-2-1
db3*l*6	1878.92	0.968	1228.55	0.999	0.104	2139.52	0.961	1442.53	0.977	0.111	28-2-1
db3*l*7	1995.40	0.964	1380.38	1.006	0.111	2562.44	0.944	1659.31	1.002	0.133	32-2-1
db4*l*1	1847.05	0.967	1251.28	0.995	0.102	2177.37	0.960	1459.10	0.983	0.113	8-2-1
db4*l*2	1557.11	0.978	1078.93	1.005	0.086	1824.73	0.972	1218.62	1.000	0.094	12-2-1
db4*l*3	1527.76	0.979	1055.87	1.004	0.085	**1656.87**	**0.977**	**1169.41**	**1.004**	**0.086**	**16-2-1**
db4*l*4	1531.49	0.978	1046.24	1.006	0.085	1670.28	0.976	1164.75	0.995	0.086	20-2-1
db4*l*5	1591.07	0.977	1062.33	1.000	0.088	1756.29	0.974	1162.33	0.995	0.091	24-2-1
db4*l*6	1600.79	0.976	1159.17	0.996	0.088	1905.21	0.969	1344.07	0.985	0.098	28-2-1
db4*l*7	1868.14	0.968	1285.62	1.009	0.103	2113.58	0.962	1479.36	0.986	0.109	32-2-1
db5*l*1	1842.49	0.969	1330.91	1.002	0.102	2158.83	0.961	1430.04	0.985	0.112	8-2-1
db5*l*2	1365.61	0.983	943.10	1.002	0.075	1610.09	0.978	1067.38	1.004	0.083	12-2-1
db5*l*3	1396.56	0.982	989.79	1.003	0.077	**1486.71**	**0.981**	**1038.95**	**0.996**	**0.077**	**16-2-1**
db5*l*4	1364.79	0.983	936.04	1.005	0.075	1535.75	0.980	1036.75	0.999	0.079	20-2-1
db5*l*5	1373.06	0.983	928.45	0.997	0.076	1570.69	0.979	1095.50	0.997	0.081	24-2-1
db5*l*6	1390.95	0.982	943.51	0.998	0.077	1681.02	0.976	1142.09	0.984	0.087	28-2-1
db5*l*7	1464.96	0.980	957.82	0.997	0.081	1973.75	0.967	1368.41	0.982	0.102	32-2-1

Note: RMSE and MAE are in cumec unit.

Table 4.10: Values of statistical parameters for ANN and WANN models Lead time: 1 week (Pancharatna station - Weekly data)

Model type	Training period RMSE	R^2	MAE	BIAS	S.I.	Testing period RMSE	R^2	MAE	BIAS	S.I.	Optimum ANN structure
ANN	4564.04	0.849	3163.04	1.006	0.282	5059.15	0.833	3587.53	1.019	0.309	4-4-1
db1*l*1	3665.44	0.903	2400.94	0.955	0.227	3482.27	0.921	2470.42	0.970	0.212	8-3-1
db1*l*2	3340.44	0.919	2128.13	1.013	0.207	3067.28	0.939	2248.44	1.009	0.187	12-3-1
db1*l*3	3114.99	0.930	2127.94	1.015	0.193	**2892.77**	**0.945**	**2168.33**	**1.030**	**0.176**	**16-3-1**
db1*l*4	3330.06	0.920	2215.56	0.977	0.206	3744.55	0.909	2717.51	1.001	0.228	20-3-1
db1*l*5	3437.85	0.914	2387.07	1.018	0.213	3823.46	0.905	2794.40	1.035	0.233	24-3-1
db1*l*6	3287.51	0.922	2165.78	0.989	0.203	4009.20	0.895	2701.62	0.941	0.245	28-3-1
db1*l*7	3404.31	0.916	2359.86	1.033	0.211	4559.07	0.864	3239.25	1.048	0.278	32-3-1
db2*l*1	2864.07	0.941	1985.76	1.011	0.177	3267.81	0.930	2313.87	1.024	0.199	8-2-1
db2*l*2	2382.87	0.959	1700.36	1.017	0.147	3065.18	0.939	2061.85	1.021	0.187	12-2-1
db2*l*3	2326.63	0.961	1528.78	0.998	0.144	2885.45	0.946	1718.94	1.007	0.176	16-2-1
db2*l*4	2426.48	0.957	1805.62	1.013	0.150	**2773.77**	**0.949**	**2116.93**	**1.015**	**0.169**	**20-2-1**
db2*l*5	2188.12	0.965	1559.33	1.006	0.135	3079.99	0.938	2022.60	0.998	0.188	24-2-1
db2*l*6	2453.45	0.956	1622.87	1.002	0.152	3144.52	0.936	2023.66	0.997	0.192	28-2-1
db2*l*7	2514.18	0.954	1818.09	1.011	0.155	3292.39	0.929	2229.47	0.997	0.201	32-2-1
db3*l*1	2477.69	0.955	1729.21	1.013	0.153	3511.97	0.919	2346.34	1.025	0.214	8-2-1
db3*l*2	2417.97	0.958	1650.03	0.987	0.149	2781.44	0.949	1845.41	0.998	0.169	12-2-1
db3*l*3	2091.81	0.968	1439.68	1.001	0.129	**2585.32**	**0.956**	**1791.36**	**1.001**	**0.158**	**16-2-1**
db3*l*4	2150.64	0.966	1611.31	1.010	0.133	2787.21	0.949	2015.73	1.014	0.170	20-2-1
db3*l*5	2281.39	0.962	1584.69	0.999	0.141	2837.71	0.947	2030.76	1.018	0.173	24-2-1
db3*l*6	2375.98	0.959	1666.59	1.015	0.147	2849.22	0.947	1979.25	1.023	0.174	28-2-1
db3*l*7	2226.40	0.964	1589.14	1.009	0.138	2860.77	0.947	2030.73	0.998	0.174	32-2-1
db4*l*1	2320.21	0.961	1580.33	0.996	0.143	3161.27	0.935	2132.03	0.998	0.193	8-2-1
db4*l*2	2240.45	0.964	1658.34	1.017	0.138	2993.51	0.942	2164.95	1.028	0.183	12-2-1
db4*l*3	1906.47	0.974	1238.63	1.007	0.118	2661.32	0.954	1599.98	1.017	0.162	16-2-1
db4*l*4	1823.62	0.976	1215.12	0.998	0.112	**2263.08**	**0.967**	**1433.70**	**1.012**	**0.138**	**20-2-1**
db4*l*5	1945.86	0.973	1378.48	1.003	0.120	2585.90	0.956	1773.88	1.014	0.158	24-2-1
db4*l*6	2161.35	0.966	1486.09	1.007	0.134	2765.96	0.950	1886.71	1.005	0.169	28-2-1
db4*l*7	2115.75	0.967	1471.29	1.010	0.131	2792.31	0.949	1961.70	1.004	0.170	32-2-1
db5*l*1	2373.44	0.959	1748.93	1.016	0.147	3061.11	0.939	2210.72	1.020	0.186	8-2-1
db5*l*2	2341.82	0.960	1926.12	1.027	0.145	2972.49	0.942	2383.71	1.044	0.181	12-2-1
db5*l*3	1893.13	0.974	1362.39	1.005	0.117	2534.68	0.958	1845.51	1.002	0.155	16-2-1
db5*l*4	1579.77	0.982	1128.34	1.005	0.098	**2047.69**	**0.973**	**1363.31**	**1.007**	**0.125**	**20-2-1**
db5*l*5	1759.68	0.977	1356.10	1.004	0.109	2574.06	0.957	1936.51	1.007	0.157	24-2-1
db5*l*6	1579.62	0.982	1116.46	0.999	0.098	2626.92	0.955	1875.64	0.993	0.160	28-2-1
db5*l*7	2026.50	0.970	1442.49	1.004	0.125	2804.96	0.949	2034.28	0.997	0.171	32-2-1

Note: RMSE and MAE are in cumec unit.

Table 4.11. Values of statistical parameters for ANN and WANN models
Lead time: 2 week (Pandu Station - weekly data)

Model type	Training period					Testing period					Optimum ANN structure
	RMSE	**R^2**	**MAE**	**BIAS**	**S.I.**	**RMSE**	**R^2**	**MAE**	**BIAS**	**S.I.**	
ANN	4996.09	0.772	3687.47	1.029	0.277	5535.43	0.740	3815.19	0.959	0.285	4-3-1
db1*l*1	4570.97	0.809	3213.85	1.015	0.253	5010.20	0.787	3630.94	0.985	0.258	8-2-1
db1*l*2	4229.52	0.837	2812.14	0.995	0.235	4412.98	0.835	3086.60	0.966	0.227	12-2-1
db1*l*3	3788.53	0.869	2559.43	1.019	0.210	4123.74	0.856	2723.09	0.993	0.212	16-2-1
db1*l*4	3540.86	0.885	2404.59	0.995	0.196	**4041.81**	**0.862**	**2844.18**	**0.971**	**0.208**	**20-2-1**
db1*l*5	3901.27	0.861	2713.31	0.990	0.216	4212.06	0.849	2873.39	0.974	0.217	24-2-1
db1*l*6	3853.96	0.864	2764.46	1.013	0.214	4357.96	0.839	2953.74	0.979	0.224	28-2-1
db1*l*7	3884.99	0.862	2855.52	1.002	0.215	7319.00	0.546	6175.60	0.698	0.376	32-2-1
db2*l*1	4150.74	0.843	2902.25	1.021	0.230	4944.37	0.793	3337.09	0.956	0.254	8-3-1
db2*l*2	3025.00	0.916	2045.91	0.986	0.168	4165.43	0.853	2662.51	0.966	0.214	12-3-1
db2*l*3	3109.41	0.912	2285.42	1.020	0.172	3601.79	0.890	2284.24	0.979	0.185	16-3-1
db2*l*4	2823.26	0.927	2008.83	1.008	0.157	**3539.25**	**0.894**	**2277.47**	**0.979**	**0.182**	**20-3-1**
db2*l*5	3082.04	0.913	2364.87	1.015	0.171	3578.99	0.891	2483.83	0.989	0.184	24-3-1
db2*l*6	2834.28	0.927	2060.57	1.017	0.157	3639.29	0.888	2385.57	1.003	0.187	28-3-1
db2*l*7	2852.62	0.926	2140.09	1.002	0.158	4023.95	0.863	2681.75	0.983	0.207	32-3-1
db3*l*1	3852.75	0.864	2725.29	1.003	0.214	4544.18	0.825	3034.36	0.965	0.234	8-2-1
db3*l*2	2885.74	0.924	2106.57	1.013	0.160	3561.10	0.892	2348.00	0.980	0.183	12-2-1
db3*l*3	2430.36	0.946	1679.37	0.994	0.135	3008.29	0.923	1895.19	0.980	0.155	16-2-1
db3*l*4	2414.86	0.946	1681.73	1.002	0.134	**2755.06**	**0.936**	**1871.04**	**0.992**	**0.141**	**20-2-1**
db3*l*5	2447.73	0.945	1790.54	1.032	0.136	2958.46	0.926	1983.15	1.012	0.152	24-2-1
db3*l*6	2552.06	0.940	1897.45	0.999	0.141	3090.13	0.919	2042.05	0.982	0.159	28-2-1
db3*l*7	2577.92	0.939	1936.72	1.028	0.143	3439.15	0.899	2248.78	0.989	0.177	32-2-1
db4*l*1	4121.58	0.845	3132.04	1.050	0.228	4594.86	0.821	3305.95	1.007	0.236	8-2-1
db4*l*2	2757.55	0.931	1963.60	1.004	0.153	3160.41	0.915	2075.16	0.969	0.162	12-2-1
db4*l*3	2750.20	0.931	2193.34	1.048	0.152	2847.39	0.931	2127.76	1.023	0.146	16-2-1
db4*l*4	2423.48	0.946	1725.33	1.008	0.134	**2658.16**	**0.940**	**1838.39**	**0.999**	**0.136**	**20-2-1**
db4*l*5	2635.95	0.936	1930.40	0.998	0.146	2799.10	0.934	1968.66	0.979	0.144	24-2-1
db4*l*6	2560.87	0.940	1858.09	1.006	0.142	3139.47	0.916	2221.86	0.959	0.161	28-2-1
db4*l*7	2626.45	0.937	1960.36	1.021	0.146	3625.55	0.889	2706.45	1.024	0.186	32-2-1
db5*l*1	3934.45	0.855	2944.65	1.016	0.236	4567.27	0.827	3100.34	0.975	0.232	8-2-1
db5*l*2	3035.67	0.886	2534.68	1.034	0.196	2786.34	0.840	2678.67	0.986	0.197	12-2-1
db5*l*3	2788.57	0.929	2054.09	1.009	0.154	2572.33	0.944	2149.73	0.990	0.132	16-2-1
db5*l*4	2431.01	0.946	1713.44	1.011	0.135	**2448.19**	**0.949**	**1662.32**	**1.001**	**0.126**	**20-2-1**
db5*l*5	2467.23	0.945	1894.34	1.023	0.138	2567.98	0.939	1730.56	1.045	0.139	24-2-1
db5*l*6	2500.34	0.934	1989.04	0.994	0.141	2645.56	0.924	1903.98	0.993	0.146	28-2-1
db5*l*7	2778.76	0.921	2012.34	1.045	0.158	2967.39	0.901	2167.36	1.023	0.176	32-2-1

Note: RMSE and MAE are in cumec unit.

Table 4.12. Values of statistical parameters for ANN and WANN models Lead time: 2 week (Pancharatna station - Weekly data)

Model	Training period					Testing period					Optimum
type	RMSE	R^2	MAE	BIAS	S.I.	RMSE	R^2	MAE	BIAS	S.I.	ANN structure
ANN	5597.84	0.773	3832.69	1.013	0.346	6844.01	0.695	4768.75	1.035	0.417	4-2-1
db1*l*1	5155.64	0.808	3427.12	1.028	0.319	6923.43	0.687	4606.84	1.001	0.422	8-3-1
db1*l*2	4352.76	0.863	3162.13	1.051	0.269	5850.92	0.777	4160.32	1.071	0.357	12-3-1
db1*l*3	3935.02	0.888	2711.77	0.995	0.243	**5798.48**	**0.781**	**4085.55**	**1.018**	**0.354**	**16-3-1**
db1*l*4	4099.65	0.878	2770.20	1.006	0.254	5913.22	0.772	4058.74	1.018	0.361	20-3-1
db1*l*5	4422.57	0.858	3242.95	1.032	0.274	5955.77	0.769	4119.26	1.024	0.363	24-3-1
db1*l*6	4226.03	0.871	3037.43	1.022	0.261	6092.31	0.758	4195.68	1.064	0.372	28-3-1
db1*l*7	4067.89	0.880	2754.31	1.011	0.252	6528.42	0.722	4224.87	1.025	0.398	32-3-1
db2*l*1	4780.48	0.835	3257.89	1.011	0.296	6951.86	0.685	4781.69	1.051	0.424	8-2-1
db2*l*2	3639.43	0.904	2488.45	1.002	0.225	5471.57	0.805	3745.57	1.029	0.334	12-2-1
db2*l*3	3256.55	0.923	2420.21	1.042	0.201	4694.27	0.856	3587.94	1.073	0.286	16-2-1
db2*l*4	3165.58	0.927	2234.19	1.002	0.196	**4559.19**	**0.865**	**3028.69**	**1.015**	**0.278**	**20-2-1**
db2*l*5	3333.45	0.919	2445.29	1.037	0.206	4693.73	0.856	3267.80	1.054	0.286	24-2-1
db2*l*6	3860.11	0.892	2653.43	0.957	0.239	4719.78	0.855	3150.31	0.947	0.288	28-2-1
db2*l*7	3417.12	0.915	2417.18	1.008	0.211	4869.39	0.845	3283.87	0.991	0.297	32-2-1
db3*l*1	4643.64	0.844	3449.27	1.042	0.287	6816.09	0.697	4915.57	1.082	0.416	8-2-1
db3*l*2	2918.18	0.938	2173.35	1.007	0.181	4000.43	0.896	2737.49	1.032	0.244	12-2-1
db3*l*3	3193.21	0.926	2449.83	1.038	0.197	3757.68	0.908	2953.92	1.060	0.229	16-2-1
db3*l*4	3003.69	0.935	2279.78	1.020	0.185	**3331.07**	**0.928**	**2530.66**	**1.036**	**0.203**	**20-2-1**
db3*l*5	2926.94	0.938	2041.07	0.997	0.181	3420.48	0.924	2493.86	1.001	0.209	24-2-1
db3*l*6	3060.02	0.932	2200.38	1.001	0.189	3640.13	0.914	2702.76	0.993	0.222	28-2-1
db3*l*7	2995.27	0.935	2094.06	0.996	0.185	3802.77	0.906	2540.28	0.956	0.232	32-2-1
db4*l*1	4432.53	0.858	3046.84	1.012	0.274	5436.34	0.807	3889.85	1.029	0.332	8-2-1
db4*l*2	2980.24	0.936	2072.48	1.012	0.184	3837.26	0.904	2620.10	1.017	0.234	12-2-1
db4*l*3	2977.87	0.936	2178.06	1.016	0.184	3570.31	0.917	2575.71	1.018	0.218	16-2-1
db4*l*4	2693.94	0.947	1862.11	1.008	0.167	**3275.13**	**0.930**	**2426.48**	**1.019**	**0.199**	**20-2-1**
db4*l*5	2859.44	0.941	2037.57	1.003	0.177	3281.89	0.929	2348.11	1.012	0.200	24-2-1
db4*l*6	2892.44	0.939	1894.86	1.017	0.179	3544.91	0.918	2499.33	1.069	0.216	28-2-1
db4*l*7	2910.33	0.939	2055.16	1.004	0.180	3630.26	0.914	2645.78	0.984	0.221	32-2-1
db5*l*1	4446.69	0.857	2986.05	0.984	0.275	5423.41	0.808	3810.81	1.021	0.331	8-2-1
db5*l*2	3026.06	0.934	2283.82	1.001	0.187	3437.66	0.923	2603.45	1.018	0.209	12-2-1
db5*l*3	2669.86	0.948	1770.43	1.006	0.165	3165.24	0.935	2050.19	1.011	0.193	16-2-1
db5*l*4	2656.32	0.949	1793.89	0.998	0.164	3117.17	0.937	2052.20	0.991	0.190	**20-2-1**
db5*l*5	2906.63	0.939	2076.11	1.017	0.179	3436.19	0.923	2476.38	1.027	0.209	24-2-1
db5*l*6	2837.08	0.942	1960.39	1.010	0.175	3592.71	0.916	2550.67	1.021	0.219	28-2-1
db5*l*7	3038.90	0.933	2378.36	1.071	0.188	3748.39	0.908	2918.64	1.093	0.228	32-2-1

Note: RMSE and MAE are in cumec unit.

4.3.1 ANN Model Results

It can be seen from **Table 4.9 and 4.11** that for ANN model for Pandu station in training and testing period the values of determination coefficient (R^2), root mean squared error (RMSE), mean absolute error (MAE), change with respect to lead time forecast. In testing period the R^2 values were found to decrease from 0.791 for 1 week lead time to 0.740 for 2 week lead time. The RMSE increases from 4971.35 cumec for 1 week lead time to 5535.43 cumec for 2 week lead time. Also, MAE increases from 3229.57 to 3815.19 cumec for 1 week and 2 week lead times, respectively. BIAS values are increasing with lead time (1 week lead time: 0.936; 2 week lead time:

0.959), indicating that the model's underestimating capacity is decreasing. Also, scatter index is increasing from 0.258 for 1 week lead time to 0.285 for 2 week lead time, indicating that model's precision is decreasing. While, during training period, the R^2 values were found to decrease from 0.835 for 1 week lead time to 0.772 for 2 week lead time. RMSE increases from 4253.35 cumec for 1 week lead time to 4996.09 cumec for 2 week lead time. Also, MAE values increased from 2802.36 to 3687.47 cumec for 1 week and 2 week lead times, respectively. BIAS values for 1 week and 2 week lead times are found to be 0.974 (representing underestimation) and 1.029 (representing overestimation), respectively. Also, scatter index is increasing from 0.236 for 1 week lead time to 0.277 for 2 week lead time, indicating that the model's precision is decreasing. The optimum ANN structures were 4-9-1 and 4-3-1 for 1 week and 2 week lead times, respectively.

Similar trend is observed for Pancharatna station. For ANN models it can be seen from **Table 4.10 and 4.12,** in testing period, the R^2 values were found to decrease from 0.833 for 1 week lead time to 0.695 for 2 week lead time. The RMSE increases from 5059.15 cumec for 1 week lead time to 6844.01 cumec for 2 week lead time. Also, MAE increases from 3587.53 to 4768.75 cumec for 1 week and 2 week lead times, respectively. BIAS values are increasing with lead time (1 week lead time: 1.019; 2 week lead time: 1.035), indicating that the model's overestimation capacity is increasing. Also, scatter index is increasing from 0.309 for 1 week lead time to 0.417 for 2 week lead time, indicating that model's precision is decreasing. While, during training period, the R^2 values were found to decrease from 0.849 for 1 week lead time to 0.773 for 2 week lead time. RMSE increases from 4564.04 cumec for 1 week lead time to 5597.84 cumec for 2 week lead time. Also, MAE values increased from 3163.04 to 3832.69 cumec for 1 week and 2 week lead times, respectively. BIAS values for 1 week and 2 week lead times are found to be 1.006 (representing overestimation) and 1.013 (representing overestimation), respectively. Also, scatter index is increasing from 0.282 for 1 week lead time to 0.346 for 2 week lead time, indicating that the model's precision is decreasing. The optimum ANN structures were 4-4-1 and 4-2-1 for 1 week and 2 week lead times, respectively. In comparison with Pancharatna, Pandu station results are better as observed flow data at Pancharatna station is highly non-stationary.

4.3.2 WANN Model Results

Similar to daily time series data models, the normalized observed data was decomposed using Daubechies wavelets of order 1 (db1) to 5 (db5) upto 7^{th} level decomposition, which were fed as input to ANN, making the model as WANN(db_i), where *i* is the order of Daubechies wavelet. The number of neurons in hidden layer was computed by trial-and-error method and performances of these are tabulated in **Table 4.9 – 4.12.**

Here also similar results are obtained as those of daily time series data models. Results of **Table 4.9 – 4.12** reveal that the performance of wavelet based hybrid WANN model is much better than the regular ANN model in testing period. For Pandu station (during testing period; **Table 4.9 and 4.11**), from the analysis it was found that the value of R^2 decreased from 0.981 (db5*l*3 – WANN(db5)) for 1 week lead time to 0.949 (db5*l*4 – WANN(db5)) for 2 week lead time, while for ANN model this decrease was from 0.791 to 0.740. Also, the RMSE values increased from 1486.71 (db5*l*3) to 2448.19 (db5*l*4) cumec for 1 week and 2 week lead time, respectively, while for ANN model this increase was from 4971.35 to 5535.43 cumec. The MAE values were found to increase from 1038.95 (db5*l*3) to 1662.32 (db5*l*4) cumec for 1 week and 2 week lead time, respectively, while for ANN model this increase was from 3229.57 to 3815.19 cumec. The scatter index was found to increase from 0.077 (db5*l*3) to 0.126 (db5*l*4) for 1 week and 2 week lead time, respectively, while for ANN model this increase was from 0.258 to 0.285. BIAS values for 1 week and 2 week lead times are found to be 0.996 (representing underestimation) and 1.001 (representing overestimation), respectively. For Pandu station (during training period; **Table 4.9 and 4.11**), from the analysis it was found that the value of R^2 decreased from 0.982 (db5*l*3 – WANN(db5)) for 1 week lead time to 0.946 (db5*l*4 – WANN(db5)) for 2 week lead time, while for ANN model this decrease was from 0.835 to 0.772. Also, the RMSE values increased from 1396.56 (db5*l*3) to 2431.01 (db5*l*4) cumec for 1 week and 2 week lead time, respectively, while for ANN model this increase was from 4253.35 to 4996.09 cumec. The MAE values were found to increase from 989.79 (db5*l*3) to 1713.44 (db5*l*4) cumec for 1 week and 2 week lead time, respectively, while for ANN model this increase was from 2802.36 to 3687.47 cumec. The scatter index was found to increase from 0.077 (db5*l*3) to 0.135 (db5*l*4) for 1 week and 2 week lead time, respectively, while for ANN model this increase was from 0.236 to 0.277. BIAS values are found to be very close to one for both the lead times. It was found that for all WANN models, during training period BIAS values are slightly greater than 1 (indicating overestimation), while these are slightly less than 1 (indicating underestimation) during testing period of Pandu station.

For Pancharatna station (during testing period; **Table 4.10 and 4.12**), from the analysis, it was found that the value of R^2 decreased from 0.973 (db5*l*4 – WANN(db5)) for 1 week lead time to 0.937 (db5*l*4 – WANN(db5)) for 2 week lead time, while for ANN model this decrease was from 0.833 to 0.695. Also, the RMSE values increased from 2047.69 (db5*l*4) to 3117.17 (db5*l*4) cumec for 1 week and 2 week lead time, respectively, while for ANN model this increase was from 5059.15 to 6844.01 cumec. The MAE values were found to increase from 1363.31 (db5*l*4) to 2052.20 (db5*l*4) cumec for 1 week and 2 week lead time, respectively, while for ANN model this increase was from 3587.53 to 4768.75 cumec. The scatter index was found to increase from 0.125 (db5*l*4) to 0.190 (db5*l*4) for 1 week and 2 week lead time, respectively, while for ANN model this increase was from 0.309 to 0.417. For Pancharatna station (during training period; **Table 4.10 and 4.12**), from the analysis, it was found that the value of R^2 decreased from 0.982 (db5*l*4 – WANN(db5)) for 1 week lead time to 0.949 (db5*l*4 – WANN(db5)) for 2 week lead time, while for ANN model this decrease was from 0.849 to 0.773. Also, the RMSE values increased from 1579.77 (db5*l*4) to 2656.32 (db5*l*4) cumec for 1 week and 2 week lead time, respectively, while for ANN model this increase was from 4564.04 to 5597.84 cumec. The MAE values were found to increase from 1128.34 (db5*l*4) to 1793.89 (db5*l*4) cumec for 1 week and 2 week lead time, respectively, while for ANN model this increase was from 3163.04 to 3832.69 cumec. The scatter index was found to increase from 0.098 (db5*l*4) to 0.164 (db5*l*4) for 1 week and 2 week lead time, respectively, while for ANN model this increase was from 0.282 to 0.346. It was found that for all WANN models, during training and testing period BIAS values are slightly greater than 1 (indicating overestimation).

The above discussion and study of **Tables 4.9 – 4.12** reveal that, in comparison with regular ANN model all WANN models have given better results for both the lead times. The WANN model was found more accurate because wavelet transform decompose the non-stationary time series data into several stationary approximation and detail time series.

4.3.3 Comparison among WANN Models with Different Daubechies Wavelets

The results obtained by WANN models using db1 to db5 mother wavelets are compared. Results of **Tables 4.9 – 4.12** depict that WANN model with db5 mother wavelet [WANN(db5)] has shown better performance compared to WANN(db1) to WANN(db4) models for both the stations.

For 1 week lead time for Pandu station in testing period (**Table 4.9**), from the analysis, it was found that the value of R^2 increased from 0.914 for db1*l*4 – WANN(db1) model to 0.981 for db5*l*3 – WANN(db5) model, while for Pancharatna station (**Table 4.10)** this increase was from 0.945 (db1*l*3) to 0.973 (db5*l*4). Also, for Pandu station the RMSE values decreased from 3191.31 for db1*l*4 – WANN(db1) model to 1486.71 cumec for db5*l*3 – WANN(db5), while for Pancharatna station this decrease was from 2892.77 (db1*l*3) to 2047.69 (db5*l*4) cumec. For Pandu station the MAE values were found to decrease from 2204.58 (db1*l*4) to 1038.95 (db5*l*3) cumec, while for Pancharatna station this decrease was from 2168.33 (db1*l*3) to 1363.31 (db5*l*4) cumec. For Pandu station the scatter index was found to decrease from 0.165 (db1*l*4) to 0.077 (db5*l*3), while for Pancharatna station this decrease was from 0.176 (db1*l*3) to 0.125 (db5*l*4). Similar results were obtained for 2 week lead time for both the stations (**Table 4.11 and 4.12**).

For 1 week lead time for Pandu station in training period (**Table 4.9**), from the analysis, it was found that the value of R^2 increased from 0.916 for db1*l*4 – WANN(db1) model to 0.982 for db5*l*3 – WANN(db5) model, while for Pancharatna station (**Table 4.10)** this increase was from 0.930 (db1*l*3) to 0.982 (db5*l*4). Also, for Pandu station the RMSE values decreased from 3032.01 for db1*l*4 – WANN(db1) model to 1396.56 cumec for db5*l*3 – WANN(db5), while for Pancharatna station this decrease was from 3114.99 (db1*l*3) to 1579.77 (db5*l*4) cumec. For Pandu station the MAE values were found to decrease from 2136.42 (db1*l*4) to 989.79 (db5*l*3) cumec, while for Pancharatna station this decrease was from 2127.94 (db1*l*3) to 1128.34 (db5*l*4) cumec. For Pandu station the scatter index was found to decrease from 0.168 (db1*l*4) to 0.077 (db5*l*3), while for Pancharatna station this decrease was from 0.193

(db1*l*3) to 0.098 (db5*l*4). Similar results were obtained for 2 week lead time for both the stations (**Table 4.11 and 4.12**).

The above discussion and the careful study of **Tables 4.9 to 4.12** reveal that, WANN model's forecasting performance increased with increase in wavelet order, giving best results for db5 mother wavelet for both the stations. **Figure 4.19** shows variation of RMSE and MAE with Daubechies wavelet order for lead time 2 day (weekly data). **Figure 4.19** depicts that, with increase in wavelet order number, the model's forecasting performance was increased, with best performance at 5th order.

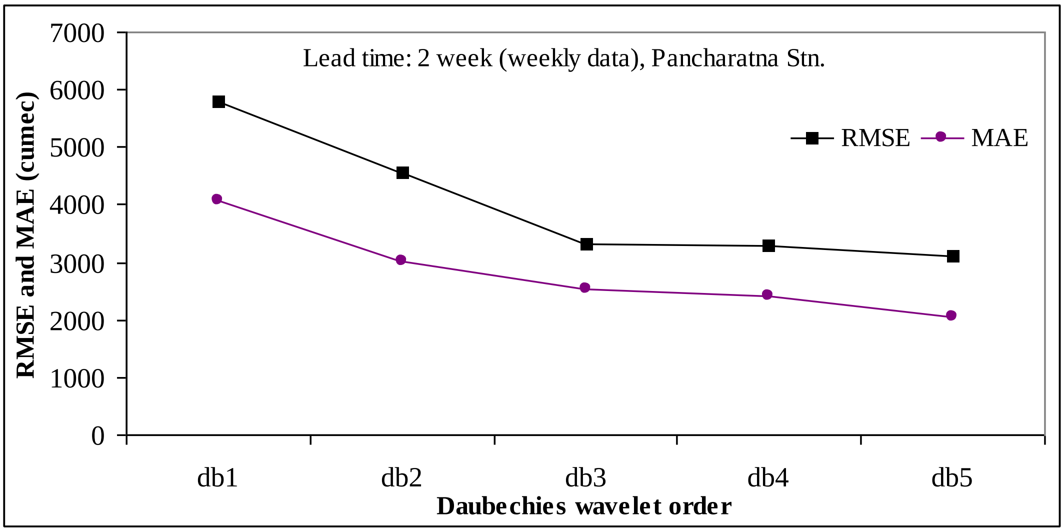


Fig.4.19. Variation of RMSE & MAE with Daubechies wavelet order in testing period for lead time 2 week (weekly data), Pancharatna Stn.

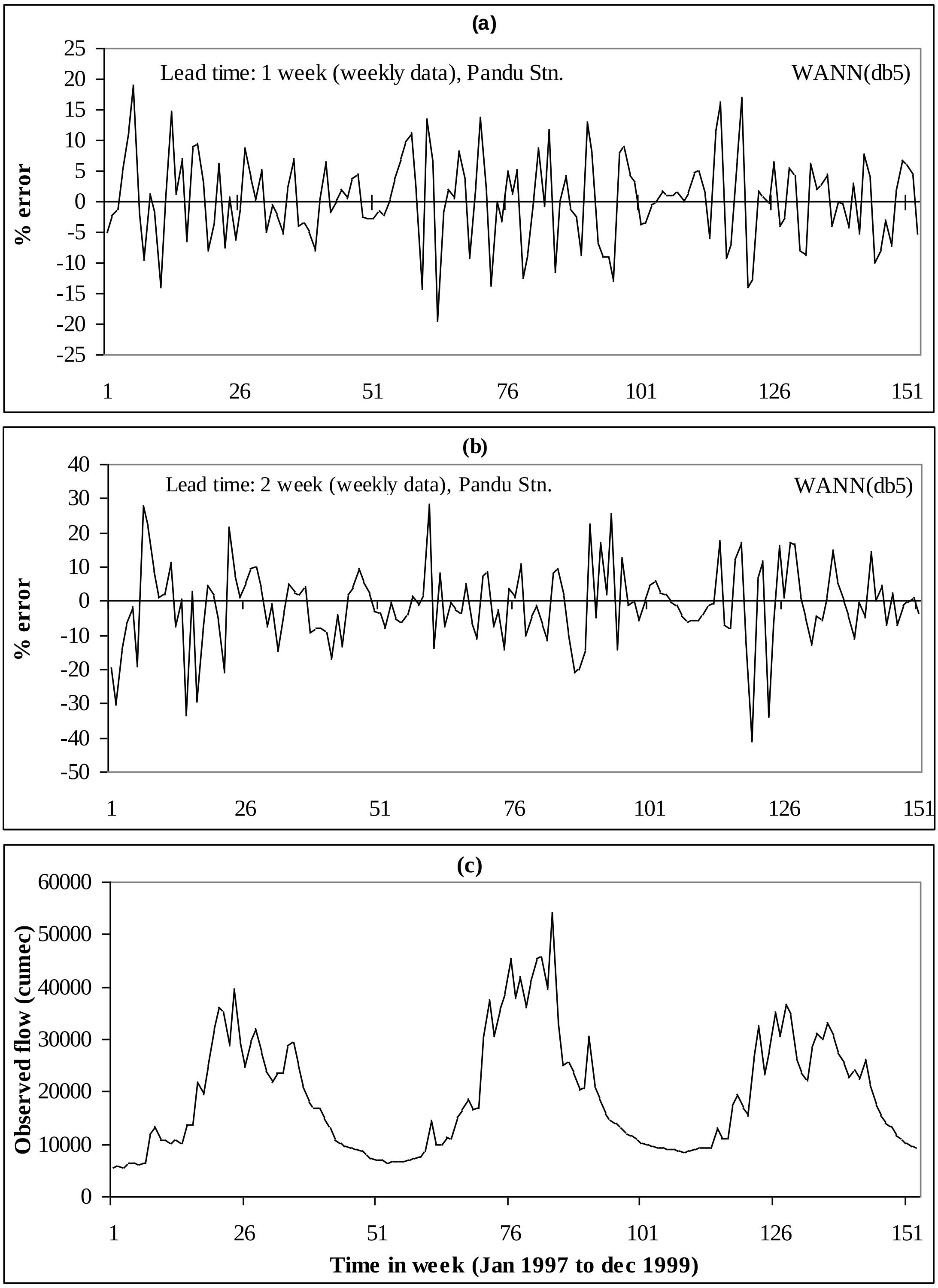


Fig.4.20. Percentage error distribution plots for (a) 1 week lead time (b) 2 week lead time, along with (c) observed flow during testing period for Pandu Stn.

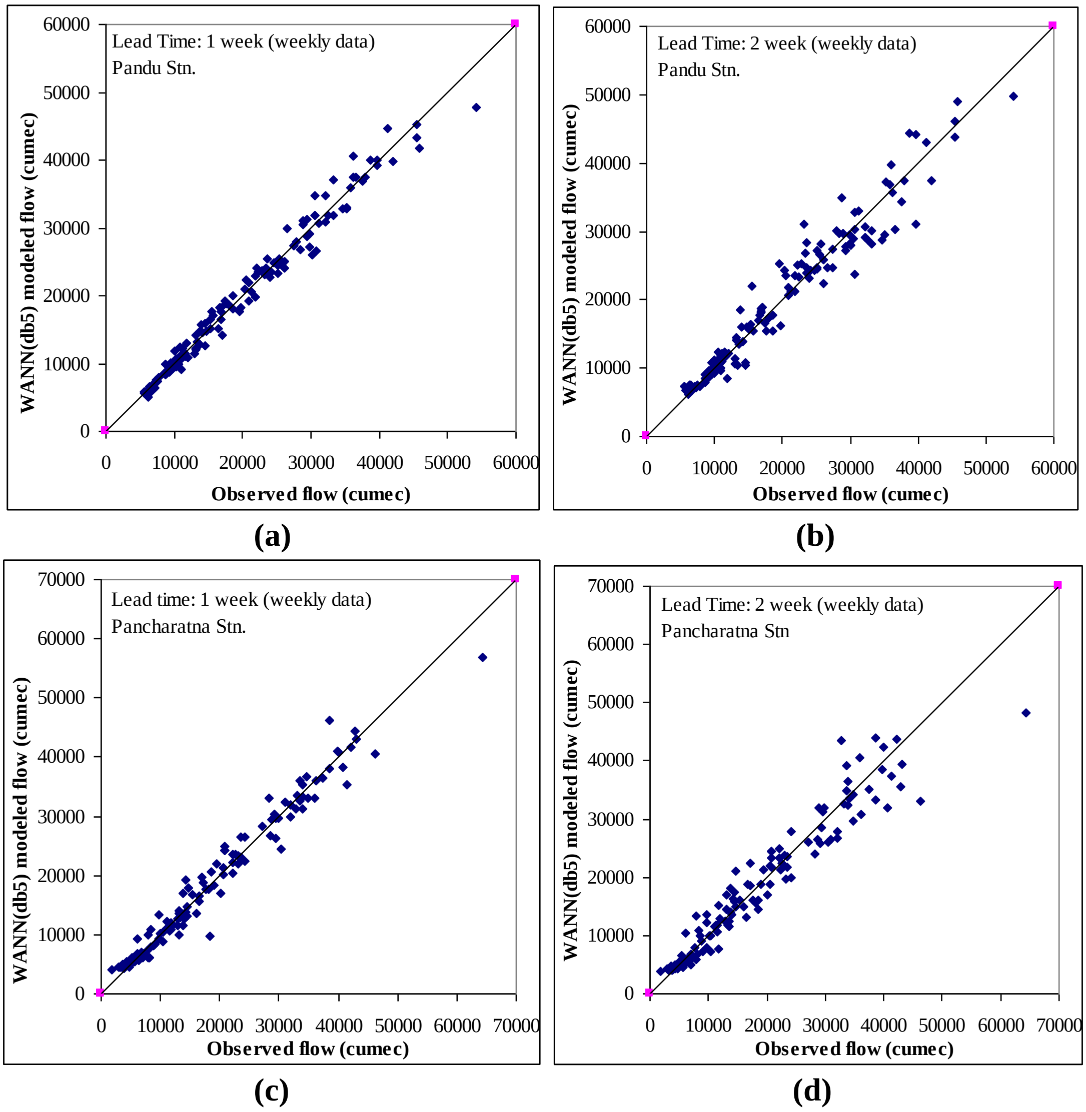


Fig. 4.21. Scatter plots for weekly data in testing period for lead times (a) 1 week (Pandu Stn.), (b) 2 week (Pandu Stn.), (c) 1 week (Pancharatna Stn.), (d) 2 week (Pancharatna Stn.)

Percentage error plots for 1 and 2 week lead times for Pandu station during testing period are shown in **Fig. 4.20 (a) and (b),** respectively. From **Figure. 4.20 (a)**, it is seen that for 1 week lead time the percentage error in peak flow computation was 11.65 % (underestimation), while the maximum underestimation error was limited to 18.94 %. On the other hand, the error in low flow computation was – 5.11 % (overestimation), while the maximum overestimation error was limited to – 19.58 %. **Figure 4.20 (b)** reveals that, for 2 week lead time the percentage error in peak flow computation was 8.06 % (underestimation), while the maximum underestimation error was to 28.44 %. On the other hand, the error in low flow computation was – 30.25 % (overestimation), while the maximum overestimation error was – 41.03 %. **Figure 4.21** shows scatter plots for 1 and 2 week lead times for Pandu and Pancharatna stations during testing period. From these plots it was observed that the results are very satisfactory (For Pandu station: R^2 for 1 and 2 week lead times 0.981 (SI = 0.077) and 0.949 (SI = 0.126), respectively; for Pancharatna station: R^2 for 1 and 2 week lead times 0.973 (SI = 0.125) and 0.937 (SI = 0.190), respectively.

4.4 MODEL RESULTS FOR MONTHLY TIME SERIES DATA

Optimum results (with low RMSE) of model runs carried out by varying order of Daubechies wavelet and decomposition level for monthly time series data for lead time 1 month for both the stations are shown in following tables (**Table 4.13 to 4.14**). Results of ANN models are also shown in the same tables for comparison.

Table 4.13. Values of statistical parameters for ANN and WANN models Lead time: 1month (Pandu station - Monthly data)

Model type	Training period					Testing period					Optimum ANN structure
	RMSE	R^2	MAE	BIAS	S.I.	RMSE	R^2	MAE	BIAS	S.I.	
ANN	4092.69	0.835	3100.18	1.025	0.226	5942.27	0.648	4804.58	0.982	0.299	4-2-1
db1*l*1	3825.52	0.856	2761.39	1.032	0.211	4819.61	0.768	3868.79	1.004	0.242	8-2-1
db1*l*2	2976.74	0.913	1935.74	1.004	0.164	**4612.62**	**0.788**	**3501.61**	**0.995**	**0.232**	**12-2-1**
db1*l*3	3438.54	0.884	2732.79	1.019	0.189	4953.00	0.755	3766.09	0.987	0.249	16-2-1
db1*l*4	4005.19	0.842	3286.64	1.049	0.221	5006.10	0.750	3966.60	0.957	0.252	20-2-1
db1*l*5	3459.63	0.882	2272.74	0.984	0.191	5647.99	0.682	4446.82	0.917	0.284	24-2-1
db2*l*1	2396.07	0.943	1838.47	0.999	0.132	3834.54	0.853	2944.95	0.979	0.192	8-2-1
db2*l*2	2787.42	0.923	2314.14	1.014	0.153	3545.40	0.875	2949.42	0.979	0.178	12-2-1
db2*l*3	2255.61	0.950	1638.45	0.970	0.124	**3309.37**	**0.891**	**2558.16**	**0.942**	**0.166**	**16-2-1**
db2*l*4	3011.72	0.911	2159.10	1.028	0.166	4601.05	0.789	3691.55	1.019	0.231	20-2-1
db2*l*5	2627.61	0.932	1842.99	0.997	0.145	5116.95	0.739	3686.47	1.009	0.257	24-2-1
db3*l*1	2215.33	0.952	1812.16	1.026	0.122	3643.44	0.868	2622.40	1.003	0.183	8-2-1
db3*l*2	1796.17	0.968	1354.25	1.012	0.099	3341.96	0.888	2515.15	1.022	0.168	12-2-1
db3*l*3	1899.32	0.964	1410.12	0.979	0.105	**3176.61**	**0.899**	**2764.69**	**0.977**	**0.159**	**16-2-1**
db3*l*4	2834.28	0.921	2075.39	1.029	0.156	3602.65	0.871	2863.46	0.965	0.181	20-2-1
db3*l*5	2954.30	0.914	2132.95	0.999	0.163	4750.27	0.775	3606.75	1.097	0.239	24-2-1
db4*l*1	2251.69	0.950	1762.24	0.973	0.124	3379.71	0.886	2637.40	0.935	0.170	8-2-1
db4*l*2	2435.30	0.942	1999.69	0.995	0.134	3194.07	0.898	2526.04	0.968	0.160	12-2-1
db4*l*3	2474.45	0.939	1799.88	0.989	0.136	**2838.53**	**0.920**	**2116.38**	**0.981**	**0.143**	**16-2-1**
db4*l*4	2898.16	0.917	2101.88	0.946	0.160	4033.05	0.838	2944.99	0.937	0.203	20-2-1
db4*l*5	2760.42	0.925	1988.09	0.984	0.152	4083.46	0.834	2762.07	1.036	0.205	24-2-1
db5*l*1	2139.61	0.955	1681.68	1.001	0.118	3218.93	0.897	2482.58	0.957	0.162	8-2-1
db5*l*2	2065.82	0.958	1636.31	0.979	0.114	**2502.65**	**0.938**	**1930.91**	**0.959**	**0.126**	**12-2-1**
db5*l*3	2120.64	0.956	1566.37	0.989	0.117	2583.82	0.933	1995.68	0.972	0.130	16-2-1
db5*l*4	1924.83	0.963	1434.88	1.007	0.106	3344.90	0.888	2603.41	0.967	0.168	20-2-1
db5*l*5	2543.85	0.936	1680.14	0.989	0.140	3794.54	0.856	3078.67	0.941	0.191	24-2-1

Note: RMSE and MAE are in cumec unit.

Table 4.14. Values of statistical parameters for ANN and WANN models Lead time: 1month (Pancharatna station - Monthly data)

Model type	Training period					Testing period					Optimum ANN structure
	RMSE	R^2	MAE	BIAS	S.I.	RMSE	R^2	MAE	BIAS	S.I.	
ANN	4855.32	0.812	3741.13	1.059	0.308	6398.43	0.690	5065.36	0.994	0.373	4-2-1
db1*l*1	3099.66	0.923	2163.24	1.003	0.197	6109.32	0.717	4088.52	1.018	0.356	8-2-1
db1*l*2	3339.61	0.911	2609.88	1.016	0.212	5643.55	0.759	4150.72	0.958	0.329	12-2-1
db1*l*3	2890.23	0.933	1793.16	1.016	0.183	**5501.73**	**0.771**	**3364.14**	**0.996**	**0.321**	**16-2-1**
db1*l*4	3168.29	0.920	2408.29	0.988	0.201	5653.25	0.758	3959.87	0.962	0.329	20-2-1
db1*l*5	3629.70	0.895	2241.10	0.969	0.230	5828.54	0.743	4084.68	0.979	0.339	24-2-1
db2*l*1	3336.49	0.908	1877.12	0.965	0.214	5873.79	0.739	3986.49	0.897	0.342	8-2-1
db2*l*2	3026.41	0.927	1966.99	0.980	0.192	5443.51	0.775	3890.61	0.915	0.317	12-2-1
db2*l*3	2405.35	0.954	1592.07	1.003	0.152	**4897.14**	**0.818**	**3530.13**	**0.901**	**0.285**	**16-2-1**
db2*l*4	3229.63	0.917	2247.89	1.071	0.205	5086.39	0.804	3863.59	1.025	0.296	20-2-1
db2*l*5	4739.49	0.821	3857.83	1.125	0.300	6505.72	0.680	5380.60	1.045	0.379	24-2-1
db3*l*1	2082.27	0.965	1510.68	0.991	0.132	4161.20	0.869	2734.23	0.957	0.243	8-2-1
db3*l*2	2221.61	0.961	1563.50	1.021	0.141	3782.85	0.892	2869.75	0.987	0.221	12-2-1
db3*l*3	2398.84	0.954	1867.36	1.013	0.152	**3552.85**	**0.904**	**2595.75**	**0.975**	**0.207**	**16-2-1**
db3*l*4	3212.76	0.918	2395.75	0.958	0.204	4424.21	0.852	3307.03	0.968	0.258	20-2-1
db3*l*5	2418.79	0.953	2017.51	1.034	0.153	5705.60	0.754	4259.40	0.948	0.334	24-2-1
db4*l*1	2098.72	0.965	1615.09	0.995	0.133	3896.19	0.885	3198.18	0.962	0.227	8-2-1
db4*l*2	1926.51	0.970	1446.06	1.007	0.122	**3065.18**	**0.929**	**2542.75**	**1.001**	**0.179**	**12-2-1**
db4*l*3	1838.91	0.973	1207.52	0.987	0.117	3251.17	0.920	2317.69	0.957	0.189	16-2-1
db4*l*4	3177.52	0.919	2049.65	0.964	0.201	3892.31	0.885	3063.06	0.906	0.227	20-2-1
db4*l*5	2984.58	0.929	1990.77	1.020	0.189	6040.68	0.724	4369.89	1.059	0.352	24-2-1
db5*l*1	2366.59	0.955	1755.19	0.973	0.150	3309.96	0.917	2581.24	0.925	0.193	8-2-1
db5*l*2	1922.55	0.970	1420.10	1.024	0.122	**<u>2733.39</u>**	**<u>0.943</u>**	**<u>2082.22</u>**	**<u>0.963</u>**	**<u>0.159</u>**	**12-2-1**
db5*l*3	1491.71	0.982	1077.16	1.010	0.094	2964.21	0.933	2244.17	0.986	0.173	16-2-1
db5*l*4	2360.02	0.956	1810.36	1.005	0.149	3683.25	0.897	3031.05	0.959	0.215	20-2-1
db5*l*5	3059.94	0.925	2277.62	1.069	0.194	5218.04	0.794	4070.13	0.998	0.304	24-2-1

Note: RMSE and MAE are in cumec unit.

4.4.1 ANN Model Results

It can be seen from **Table 4.13** that for optimum ANN model for Pandu station during training period the values of determination coefficient (R^2), root mean squared error (RMSE), mean absolute error (MAE) are 0.835, 4092.69 cumec, and 3100.18, respectively, while during testing period these values are 0.648, 5942.27 cumec, and 4804.58 cumec, respectively. On the other hand, for Pancharatna station (**Table 4.14**) during training period the values of R^2, RMSE, and MAE are 0.818, 4855.32 cumec, and 3741.13 cumec, respectively, while during testing period these values are 0.690, 6398.43 cumec, and 5065.36 cumec, respectively. During testing period, as the R^2 values are very low representing a very weak correlation, ANN model can not be applied for monthly river flow prediction.

4.4.2 WANN Model Results

Similar to daily and weekly time series data models, the normalized observed data was decomposed using Daubechies wavelets of order 1 (db1) to 5 (db5) upto 5^{th} level, which were fed as input to ANN. The number of neurons in hidden layer was computed by trial-and-error method and performances of these are tabulated in **Table 4.13 and 4.14.**

Here also similar results are obtained as those of daily and weekly time series data models. It can be seen from **Table 4.13** that for Pandu station during testing period, db5*l*2 – WANN(db5) model have shown best performance. The values of determination coefficient (R^2), root mean squared error (RMSE), mean absolute error (MAE), BIAS and S.I. are 0.938, 2502.65 cumec, 1930.91 cumec, 0.959, and 0.126, respectively. Results of **Table 4.14** reveal that, for Pancharatna station during testing period, db5*l*2 – WANN(db5) model have shown best performance. The values of determination coefficient (R^2), root mean squared error (RMSE), mean absolute error (MAE), BIAS, and S.I. are 0.943, 2733.39 cumec, 2082.22 cumec, 0.963, and 0.159, respectively. Also, it can be seen from **Table 4.13** that for Pandu station during training period, (db5*l*2 – WANN(db5) model the values of determination coefficient (R^2), root mean squared error (RMSE), mean absolute error (MAE), BIAS and S.I. are 0.958, 2065.82 cumec, and 1636.31 cumec, 0.979, and 0.114, respectively. Results of **Table 4.14** reveal that, for Pancharatna station during training period, (db5*l*2 – WANN(db5) model), the values of determination coefficient (R^2), root mean squared error (RMSE), mean absolute error (MAE), BIAS and S.I. are 0.970, 1922.55 cumec, and 1420.10 cumec, 1.024, and 0.122, respectively.

From the above discussion it is worth to mention that, in comparison with regular ANN model WANN model has given better results. The WANN model was found more accurate because wavelet transform decomposes the non-stationary time series data into several stationary approximation and detailed time series.

4.4.3 Comparison among WANN Models with Different Daubechies Wavelets

The results obtained by WANN models using db1 to db5 mother wavelets are compared. Results of **Table 4.13 and 4.14** depicts that WANN model with db5 mother wavelet [WANN(db5)] has shown better performance compared to WANN(db1) to WANN(db4) models for both the stations.

For 1 month lead time for Pandu station in testing period (**Table 4.13**), from the analysis, it was found that the value of R^2 increased from 0.788 for db1*l*2 – WANN(db1) model to 0.938 for db5*l*2 – WANN(db5) model, while for Pancharatna station (**Table 4.14)** this increase was from 0.771 (db1*l*3) to 0.943 (db5*l*2). Also, for Pandu station the RMSE values decreased from 4612.62 (db1*l*2) to 2502.65 (db5*l*2) cumec for, while for Pancharatna station this decrease was from 5501.73 (db1*l*3) to 2733.39 (db5*l*2) cumec. For Pandu station the MAE values were found to decrease from 3501.61 (db1*l*2) to 1930.91 (db5*l*2) cumec, while for Pancharatna station this decrease was from 3364.14 (db1*l*3) to 2082.22 (db5*l*2) cumec. For Pandu station the scatter index was found to decrease from 0.232 (db1*l*2) to 0.126 (db5*l*2), while for Pancharatna station this decrease was from 0.321 (db1*l*3) to 0.159 (db5*l*2). For 1 month lead time for Pandu station in training period (**Table 4.13**), from the analysis, it was found that the value of R^2 increased from 0.913 for db1*l*2 – WANN(db1) model to 0.958 for db5*l*2 – WANN(db5) model, while for Pancharatna station (**Table 4.14)** this increase was from 0.933 (db1*l*3) to 0.970 (db5*l*2). Also, for Pandu station the RMSE values decreased from 2976.74 (db1*l*2) to 2065.82 (db5*l*2) cumec for, while for Pancharatna station this decrease was from 2890.23 (db1*l*3) to 1922.55 (db5*l*2) cumec. For Pandu station the MAE values were found to decrease from 1935.74 cumec (db1*l*2) to 1636.31 (db5*l*2) cumec, while for Pancharatna station this decrease was from 1793.16 (db1*l*3) to 1420.10 (db5*l*2) cumec. For Pandu station the scatter index was found to decrease from 0.164 (db1*l*2) to 0.114 (db5*l*2), while for Pancharatna station this decrease was from 0.183 (db1*l*3) to 0.122 (db5*l*2).

The above discussion and the careful study of **Table 4.13 and 4.14** reveal that, WANN model's forecasting performance increased with increase in wavelet order, giving best results for db5 mother wavelet for both the stations.

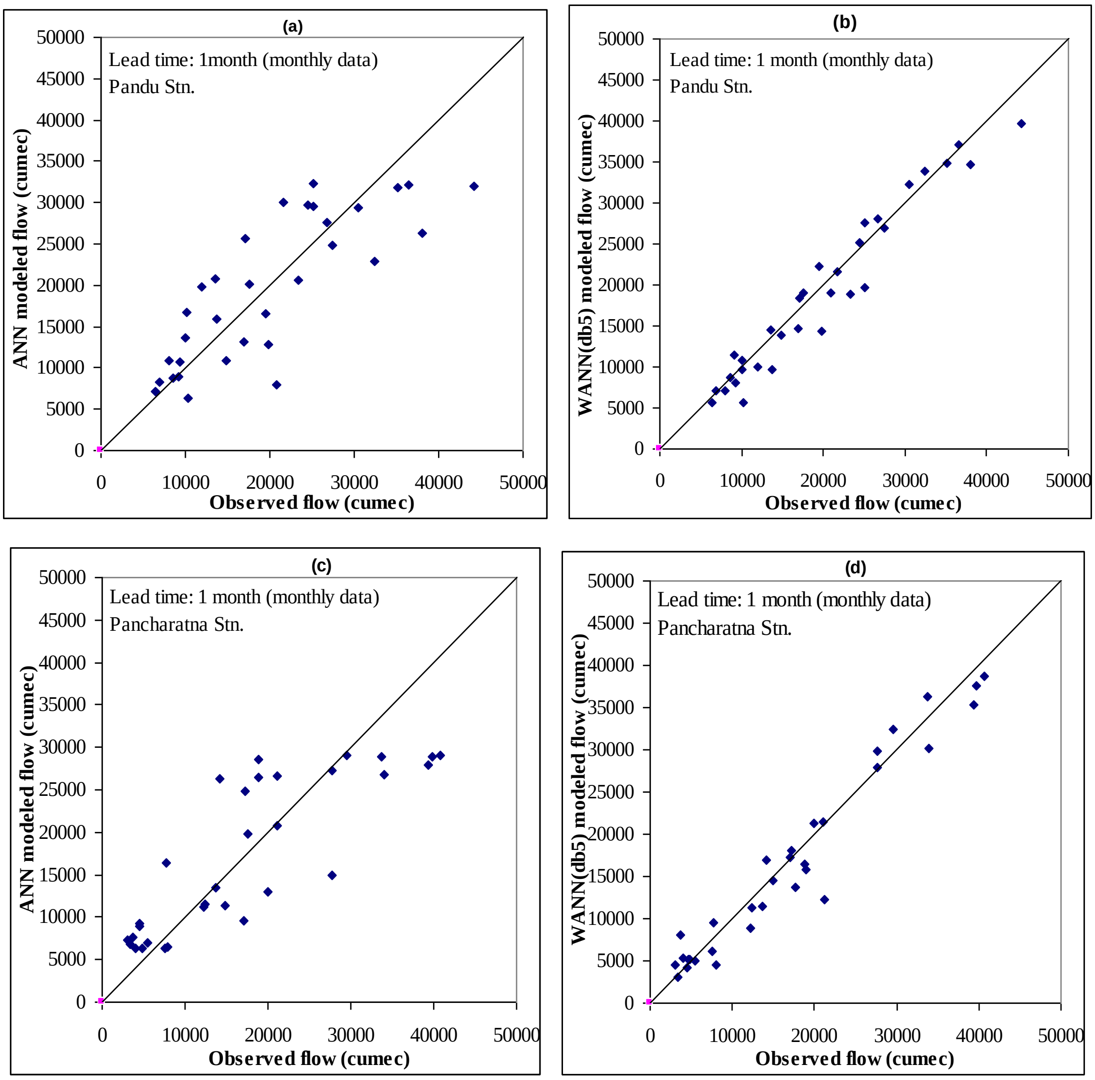


Fig.4.22. Scatter plots between observed and (a) ANN modeled (Pandu Stn.), (b) WANN(db5) modeled (Pandu Stn.), (c) ANN modeled (Pancharatna Stn.), (d) WANN(db5) modeled (Pancharatna Stn.) during testing period (lead time: 1 month)

Figure 4.22 shows scatter plots comparing ANN and WANN(db5) models for Pandu and Pancharatna stations for 1 month lead time during testing period. WANN model performed better than ANN model. WANN model captured extreme events fairly well, while ANN frequently failed to do so. Pancharatna station results are comparatively poor than Pandu due to high skewness coefficient (0.782) during testing period.

4.5 ACCURACY OF DEVELOPED MODELS IN SIMULATING HIGH FLOWS

Table 4.15 - A. Accuracy of developed models in simulating high flows in testing period - Daily time series data

Lead time	Pandu station			Pancharatna station		
(day)	Observed*	Modeled*	% error#	Observed*	Modeled*	% error#
2	54100	49333.62	8.81	76235.79	66615.93	12.62
	52800	47705.05	9.65	64284.09	67016.15	-4.25
	51100	46868.77	8.28	60988.12	63806.23	-4.62
	50600	48538.13	4.07	58692.19	62046.31	-5.71
	47700	45591.97	4.42	56462.55	52765.86	6.55
	47200	45171.92	4.30	56243.73	57142.63	-1.60
	46900	45865.04	2.21	55376.51	53610.86	3.19
	46500	45049.85	3.12	54734.83	49901.15	8.83
	46300	46207.67	0.20	52449.74	43759.40	16.57
3	54100	52483.84	2.99	76235.79	64754.69	15.06
	52800	51795.21	1.90	64284.09	67781.23	-5.44
	51100	50914.12	0.36	60988.12	63324.26	-3.83
	50600	50575.35	0.05	58692.19	56801.57	3.22
	47700	48040.72	-0.71	56462.55	50222.36	11.05
	47200	47809.55	-1.29	56243.73	60142.40	-6.93
	46900	48842.32	-4.14	55376.51	50228.62	9.30
	46500	45987.59	1.10	54734.83	48938.28	10.59
	46300	47878.10	-3.41	52449.74	42134.08	19.67
4	54100	49483.84	8.53	76235.79	59274.86	22.25
	52800	49120.59	6.97	64284.09	62978.95	2.03
	51100	48487.98	5.11	60988.12	64006.93	-4.95
	50600	48251.04	4.64	58692.19	55767.51	4.98
	47700	47741.07	-0.09	56462.55	47547.44	15.79
	47200	46230.40	2.05	56243.73	59968.51	-6.62
	46900	46412.02	1.04	55376.51	52489.83	5.21
	46500	46548.52	-0.10	54734.83	45018.34	17.75
	46300	46743.03	-0.96	52449.74	45950.70	12.39
7	54100	49388.42	8.71	76235.79	60163.91	21.08
	52800	48776.22	7.62	64284.09	61681.87	4.05
	51100	47315.04	7.41	60988.12	61424.00	-0.71
	50600	47847.18	5.44	58692.19	57412.28	2.18
	47700	45176.20	5.29	56462.55	43381.85	23.17
	47200	45855.80	2.85	56243.73	59102.61	-5.08
	46900	44612.00	4.88	55376.51	54412.53	1.74
	46500	46558.10	-0.12	54734.83	41739.93	23.74
	46300	42953.62	7.23	52449.74	41732.07	20.43
14	54100	43952.18	18.76	76235.79	50264.56	34.07
	52800	42155.52	20.16	64284.09	50710.60	21.11
	51100	40461.27	20.82	60988.12	51266.53	15.94
	50600	44026.75	12.99	58692.19	50150.72	14.55
	47700	38528.06	19.23	56462.55	40110.01	28.96
	47200	42151.15	10.70	56243.73	51756.62	7.98
	46900	41436.36	11.65	55376.51	49874.97	9.93
	46500	42969.04	7.59	54734.83	42586.03	22.20
	46300	40778.10	11.93	52449.74	46852.17	10.67

*Observed and modeled values are in cumec. # Positive – underestimation, Negative - overestimation

Table 4.15 - B. Accuracy of developed models in simulating high flows in testing period - Weekly time series data

Lead time (week)	Pandu station			Pancharatna station		
	Observed*	Modeled*	% error#	Observed*	Modeled*	% error#
1	54100	47795.86	11.65	64284.09	56867.90	11.54
	45800	41837.94	8.65	46282.53	40540.88	12.41
	45500	43278.53	4.88	43053.3	43035.99	0.04
	45500	45198.76	0.66	42901.48	44297.36	-3.25
	42000	39838.88	5.15	42155.62	41696.43	1.09
	41200	44714.79	-8.53	41431.39	35395.45	14.57
	39700	39955.11	-0.64	40728.18	38180.28	6.26
2	54100	49736.60	8.07	64284.09	48258.29	24.93
	45800	48989.89	-6.96	46282.53	33109.10	28.46
	45500	43760.21	3.82	43053.30	39420.59	8.44
	45500	46124.38	-1.37	42901.48	35573.03	17.08
	42000	37481.25	10.76	42155.62	43746.78	-3.77
	41200	42975.29	-4.31	41431.39	37353.50	9.84
	39700	44209.85	-11.36	40728.18	31935.40	21.59

*Observed and modeled values are in cumec. # Positive – underestimation, Negative – overestimation.

Table 4.15 - C. Accuracy of developed models in simulating high flows in testing period - Monthly time series data

Lead time (month)	Pandu station			Pancharatna station		
	Observed*	Modeled*	% error#	Observed*	Modeled*	% error#
1	44200	39748.03	10.07	40728.18	38740.64	4.88
	38004	34668.60	8.78	39777.78	37628.31	5.40
	36521	37107.27	-1.61	39382.23	35267.02	10.45
	35214	34799.36	1.18	33983.22	30136.92	11.32
	32456	33926.33	-4.53	33782.10	36338.47	-7.57
	30456	32238.32	-5.85	29546.58	32489.12	-9.96
	27485	26903.45	2.12	27663.36	27936.39	-0.99

*Observed and modeled values are in cumec. # Positive – underestimation, Negative – overestimation.

For investigating the ability of WANN model in simulating the peak values, first few highest observed river flows in testing period were compared with their simulated values. **Tables 4.15 A, B, and C** show relative percentage error between observed daily, weekly and monthly river flow and their best fitted simulated values. Study of these tables reveals that, the hybrid WANN model is effective for flood and drought analysis.

CHAPTER 5

SUMMARY AND CONCLUSIONS

5.1 SUMMARY OF WORK

The main purpose of the present study is to examine the applicability and generalization capability of wavelet transformation (WT) as a preprocessing technique combined with artificial neural network (ANN) using case study of forecasting Brahmaputra River flow using daily (short term), weekly (long term), and monthly (long term) time series data for multiple lead times. Ten year daily time series data for the two stations namely Pandu (u/s) and Pancharatna (d/s) located on Brahmaputra River are used in the study. Out of 10 years data, first 7 years (70 %) data is used for training and remaining 3 years (30 %) data used for testing. The above two stations were selected due to highly varied statistical parameters of observed flow at these the stations. Using daily time series data, forecasting is carried out for lead times 2, 3, 4, 7, and 14 day. Using weekly time series data forecasting is carried out for lead times of 1, and 2 week, while using monthly data forecasting is carried out for lead time of 1 month. Total seven input combinations were finalized based on auto-correlations for flow series. Then the optimal input combinations for every lead time and for every time series is finalized using trial and error procedure by varying number of neurons in hidden layer from 1 to 20 using three layer FFBP network and Lavenberg-Marquardt as training algorithm with tansig as activation function. For the selection of optimal input combinations the total number of trials taken were **1120** {7 (no. of input combinations) × 20 (no. of neurons) × 8 (no. of lead times for all time series i.e. daily, weekly, and monthly). The optimal combinations were selected based on lowest RMSE.

Then a three-layer feed forward backpropagation ANN models without data pre-processing were developed to forecast river discharge for multiple lead times for both short term and long time series data by varying the number the of neurons in hidden layer from 1 to 20. The total number of ANN models developed for Pandu station were 160 {20 (no. of neurons in hidden layer) × 8 (total no. of lead times for all time series)}. The network is trained with Lavenberg-Marquardt training algorithm and tansig as activation function.

After developing ANN models, hybrid wavelet transform-neural network (WANN) models are developed. In WANN model the original time series is decomposed into various decomposition levels (up to 7 levels) (approximation and details) using Daubechies wavelets of order 1 (db1) to 5 (db5), which is fed as input to ANN to get the output at the required lead time. For WANN model the number of neurons in hidden layer was varied from 2 to 15. For Pandu station for daily data the total number of WANN models developed are **3920** {14 (no. of neurons in hidden layer) × 7 (no. of decomposition levels) × 8 (no. of lead time) × 5 (no. of wavelets used). For Pandu station for weekly and monthly data the total number of WANN models developed are **980** (14 × 7 × 2 × 5) and **350** ((14 × 7 × 1 × 5), respectively.

For developing WANN models for Pancharatna station for daily data, same optimum ANN structure obtained for Pandu station is used. Hence the total number of WANN models developed for Pancharatna station for daily data is **280** (1 × 7 × 8 × 5). While for weekly and monthly data WANN models were developed by taking all runs similar to Pandu station. Hence for Pancharatna station for weekly and monthly data the total number of WANN models developed are **980** (14 × 7 × 2 × 5) and **350** (14 × 7 × 1 × 5), respectively. Also, the effect of decomposition level on WANN model efficiency is studied.

The models are evaluated through five performance measures: RMSE, MAE, R^2, BIAS and Scatter Index. No contradicting results are occurred with respect to above performance indices.

The data set which produces best forecast results are selected as proper data pre-processing technique for flow forecasting. The forecast results produced by optimum ANN model from each data set are compared with WANN model to select appropriate pre-processing technique.

5.2 CONCLUSIONS

Following are the important conclusions drawn from this study:

1. Use of wavelet transform as pre-processing technique is highly beneficial to improve forecasting accuracy that is revealed through the study.
2. Model with small number of input nodes and their smaller corresponding network structure produces better forecast results.
3. Adopting tansig activation function results first convergence and reduces complexity of ANN as observed in most cases of the study.
4. Selection of input parameters based on auto-correlation coefficient helps in reducing computational time.
5. In modeling higher fluctuatious of flow during monsoon period, it has been well predicted by the proposed WANN model.
6. Compared to ANN models, all WANN forecasting models with Daubechies wavelets of order 1 (db1) to 5 (db5) have better results for multiple lead times for both short term and long term time series forecasting.
7. WANN model’s forecasting performance increases with increase in wavelet order, giving best results for db5 mother wavelet for all lead times for both the stations. The db5 wavelet has a reasonable support and also has good time-frequency localization property and these together enable the model to capture both the underlying trend as well as the short term variablities in the time series better than db1 to db4 wavelet based forecast model.
8. For WANN models with db1 to db5 mother wavelets, the models efficiency increased with decomposition level up to a certain optimum level, there after it was decreased. The errors were not increased proportionately with the higher resolution as the random parts of original time series were mainly in the lower resolution level, and obviously the prediction errors were also mainly in the lower resolution level.

5.3 PROMINENT RESULTS FROM THE STUDY

- Time series data has been preprocessed using wavelet transform to produce better forecast results using case study of two u/s and d/s gauging stations.
- The proposed WANN hybrid approach, using Daubechies wavelets, reveals the high potential and ability to handle non stationary hydrologic time series.
- The forecasting performance is also influenced by proper type of wavelet as well as decomposition level.
- The time series of various temporal scales are well forecasted by proposed approach both for short term and long term forecasting along with multistep lead time consistently.
- The proposed WANN model offers wide scope and high potential in handling hydrologic time series than other models as observed in the study.

5.4 FUTURE SCOPE

- The work can be extended with other type of wavelets to explore higher accuracy aspect.
- Adaptive Neural Fuzzy Inference System (ANFIS), Fuzzy Logic (FL), Support Vector Machine (SVM) can be integrated with wavelet for further study to assess the performance under various data constraints.
- Other hydrological time series variables such as rainfall, temperature, evapotranspiration can be used as model inputs to examine the forecasting performance of proposed WANN model.
- To test the WANN models, these can be applied to sites with different hydro-climatic condition. Also uncertainty estimation can be done.

REFERENCES

Abghari, H., Ahmadi, H., Besharat, S., and Rezaverdinejad, V. (2012). “Prediction of Daily Pan Evaporation using Wavelet Neural Networks.” *Water Resources Management*, 26(12), 3639–3652.

Adamowski, J. F. (2008). “Development of a short-term river flood forecasting method for snowmelt driven floods based on wavelet and cross-wavelet analysis.” *Journal of Hydrology*, 353(3-4), 247–266.

Adamowski, K., Prokoph, A., Adamowski, J.(2009). “Development of a new method of wavelet aided trend detection & estimation.” Hydrol. Process. 23, 2686–2696.

Adamowski, J., Karen, S. (2010), “Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds.” Journal of Hydrology, 390, 85-91.

Adamowski, J. and Chan H. F. (2011). “A wavelet neural network conjunction model for ground water level forecasting.” J. of Hydrology, 407, 28-40.

Addison, P. S., Murry, K. B., Watson, J. N. (2001). “Wavelet transform analysis of open channel wake flows.” ASCE J. Engg. Mechanics, vol.127 (1), 58-70.

Adeloye, A. J., and Munari, A. D. (2006). "Artificial neural network based generilized storage-yield-reliability models using Levenberg - Marquardt algorithm." J. of Hydrology, 362, 215-230.

Asadi, S., Shahrabi, J., Abbaszadeh, P., and Tabanmehr, S. (2013). “A new hybrid artificial neural networks for rainfall–runoff process modeling.” *Neurocomputing*, Elsevier, 121, 470–480.

ASCE Task Committee, (2000b). “Artificial neural networks in hydrology-II: Hydrologic applications.” J. Hydrologic Engineering, ASCE 5(2), 124–137.

Banhatti, A. G, and Deka, P. C. (2012). “Performance evaluation of artificial neural network model using data preprocessing in non-stationary hydrologic time series.” CiiT J. of Artificial Intelligent Systems and Machine Learning, 4 (4), 223-228.

Beale, M. H., Hagan, M. T., and Demuth, H. B. (2012). MATLAB neural network toolbox user's guide. MathWorks, Inc.

Besaw, L. E., Rizzo, D. M., Bierman, P. R., Hackett, W. R. (2010), “Advances in ungauged stream flow prediction using artificial neural networks”, Journal of Hydrology 386 (2010) 27–37.

Cannas B, Fanni A. See L, Sias G. (2006). "Data preprocessing for river flow forecasting using neural networks-wavelet transforms and data partitioning." Phys Chem Earth 31(18);1164-1171.

Campisi-Pinto, S., Adamowski, J., and Oron, G. (2012). "Forecasting Urban Water Demand Via Wavelet-Denoising and Neural Network Models. Case Study: City of Syracuse, Italy." *Water Resources Management*, 26(12), 3539–3558.

Chandramouli, V., and Deka, P. (2005). "Neural network based decision support model for optimal reservoir operation." Water Resources Management, 19, 447-464.

Chen, G.Y., Xie, W.F. (2007). "Pattern recognition with SVM and dual-tree complex wavelets." Image Vision Comput. 25 (6), 960–966.

Cheng, C. T., Xie, J. X., Chau, K. W., Layeghifard, M. (2008). "A new indirect multistep- ahead prediction model for a long-term hydrologic prediction." J. Hydrol. 361, 118–130.

Chou, C. M., and Wang, R. Y. (2004). "Application of wavelet based multi-model Kalman filters to real-time flood forecasting." Hydrologic Processes,18,987-1008.

Christian, W., Dawsonand, R. (1998). "An artificial neural network approach to rainfall runoff modeling." Hydrological Sciences Journal , 43(1), 47-66.

Cigizoglu, H. K. and Kisi, O. (2005),"Flow prediction by three back propagation techniques using k-fold partitioning of neural network training data", Nordic Hydrology Vol 36 No 1 pp 49–64 q IWA Publishing.

Coulibaly, P., and Burn, H. D.(2004). "Wavelet analysis of variability in annual Canadian streamflows." Water Resources Research, 40(3), 1-14.

Daubechies, I., 1992. Ten Lectures on Wavelets. SIAM, Philadelphia, PA.

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

Deka, P., and Chandramouli, V. (2005), "Fuzzy neural network model for hydrologic flow routine", J. of Hydrologic Engineering, ASCE, 10 (4), 302–314.

Deka, P.C., Mandal, S., Prahlada, R. (2010). "Multiresulution Wavelet-ANN model for significant wave height forecasting." *Proceedings of National conference on hydraulics and water resources,* HYDRO-2010, Dec-16-18th, 230-235.

Deka, P. C., and Prahlada, R. "Discrete wavelet neural network approach in significant wave height forecasting for multistep lead time." Ocean Engineering, 43, 32-42, 2012.

Foufoula-Georgiou, E., Kumar, P., 1994. Wavelets in Geophysics. Academic Press, San Diego.

Goyal, M. K. (2013). “Monthly rainfall prediction using wavelet regression and neural network: an analysis of 1901–2002 data, Assam, India.” *Theoretical and Applied Climatology*, (2011).

Goswami, D. C. (1985). “Brahmaputra River, Assam, India: Physiography, Basin Denudation, and Channel Aggradation.” Water Resources Research, 21 (7), 959-978.

Hagan, M. T., and Menhaj, M. B. (1994). "Training feedforward networks with the Marquardt algorithm." IEEE Tras. on Neural networks, 5 (6), 989 - 993.

Haykin, S., 1998, Neural Networks – A comprehensive Foundation, 2nd ed. Prentice Hall.

Jain, S. K., and Chalisgaonkar, D. "Setting up stage-discharge relations using ANN." ASCE J. of Hydro. Engineering, 5 (4), 428-433, (2000).

Karthikeyan, L., and Nagesh Kumar, D. (2013). “Predictability of nonstationary time series using wavelet and EMD based ARMA models.” *Journal of Hydrology*, Elsevier B.V., 502, 103–119.

Karunanithi, N., Grenney, W. J., Whitly, and Bovee, K. (1994). “Neural networks for river flow prediction.” J. Computing in Civil Engg., ASCE, 8(2), 201-220.

Kim, T. and Valdes, J. B. (2003). “Nonlinear model for drought forecasting based on a conjunction of wavelet transform and neural networks.” ASCE Journal of Hydrologic Engineering, l.8 (6), 319-328.

Kisi, O. (2007 a). “Streamflow forecasting using different artificial neural network algorithms.” J. Hydrologic Engg., ASCE, 12 (5), 532-539.

Kisi. O. (2007 b). “Development of Streamflow-Suspended Sediment Rating Curve Using a Range Dependent Neural Network.” Int. J. Sci. & Tech.,

Kisi, O. (2008). “River flow forecasting and estimation using different artificial neural network techniques.” Hydrol. Res. 39 (1), 27–40. 2(1), 49-61.

Kisi, O. (2009a). “Neural network and wavelet conjunction model for modeling monthly level fluctuations in Turkey.” Hydrol. Process. 23, 2081–2092.

Kisi, O. (2009b). “Wavelet regression model as an alternative to neural networks for monthly streamflow forecasting.” Hydrol. Process. 23, 3583–3597.

Kişi, Ö. (2009c). “Daily suspended sediment estimation using neuro-wavelet models.” *International Journal of Earth Sciences*, 99(6), 1471–1482.

Kisi, O. (2009d). “Neural Networks and Wavelet Conjunction Model for Intermittent Stream flow Forecasting.” Journal of Hydrologic Engineering, 14(8), 773–782.

Kisi, O. (2010). “Daily suspended sediment estimation using neuro-wavelet models.” Int J Earth Sci (Geol Rundsch) (2010) 99:1471–1482

Kişi, O. (2011). “A combined generalized regression neural network wavelet model for monthly streamflow prediction.” *KSCE Journal of Civil Engineering*, 15(8), 1469–1479.

Kisi, O., and Shiri, J. (2011). “Precipitation Forecasting Using Wavelet-Genetic Programming and Wavelet-Neuro-Fuzzy Conjunction Models.” *Water Resources Management*, 25(13), 3135–3152.

Kisi, O., and Cimen, M. (2012). “Precipitation forecasting by using wavelet-support vector machine conjunction model.” Engineering Applications of Artificial Intelligence, 25, 783-792.

Kisi, O., Shiri, J., and Tombul, M. (2013). “Modeling rainfall-runoff process using soft computing techniques.” *Computers & Geosciences*, Elsevier, 51, 108–117.

Krishna, B. “Comparison of wavelet based ANN and Regression models for reservoir inflow forecasting”. J. of Hydrologic Engg. (ASCE) 2013. (Accepted manuscript, not copyedited).

Labat, D. (2005). “Recent advances in wavelet analyses: part 1. A review of concepts.” J. Hydrol. 314, 275–288.

Labat, D. (2008). “Wavelet analysis of the annual discharge records of the world's largest rivers.” Adv. Water Res. 31, 109–117.

Labat, D., Ababou, R., Mangin, A. (2000). “Rainfall-runoff relations for karstic springs. Part II: continuous wavelet and discrete orthogonal multiresolution analyses.” J. Hydrol. 238, 149–178.

Lam, H. K., Ling, S., and Nguyen, H. T. (2012). "Computational Intelligence and its applications: Evolutionaru computation, fuzzy logic, neural network and support vector machine techniques." World Scientific Publishing Company.

Li, X., Ding, J, Li, H. (1997). “Wavelet analysis and its potential application to hydrology and water resources.” J. Sichuan Union university (Engineering science); vol. 1(4), 49-52.

Li, X., Ding, J., Li, H. (1999). “Combing neural network models based on wavelet transform.” J. of Hydraulic, vol.2, 1-4.

Lu, R.Y. (2002). “Decomposition of interdecadal and interannual components for North China rainfall in rainy season,Chin.” J. Atmosp., 26, 611-624.

Maheswaran, R. and Khosa, R. (2012). “Comparative study of different wavelets for hydrologic forecasting.” Computers and Geosciences, 46, 284-295.

Mallat, S. (1989 a). “Multiresolution approximations and wavelet orthonormal bases of L2(R).” Trans. Am. Math. Soc. 315 (1), 69–87.

Mallat, S., G. (1989 b). “A theory for multiresolution signal decomposition: The wavelet representation.” IEEE Trans. Pattern Anal. Mach. Intell., 11(7), 674-693.

Mallat, S. G. (1998). “A wavelet tour of signal processing.” Academic, San Diego.

Meyer, Y., 1992. Wavelet and Applications. Springer, Berlin.

Mehmet, C. D., Venancio, A, Kahya, E. (2009). “Flow forecast by SWAT model and ANN in Pracana basin, Portugal.” Advances in Engineering Software, 40, (2009), 467–473.

Milly, P. C. D., J. Betancourt, M., Falkenmark, R. M., Hirsch, Z. W., Kundzewicz, D. P., Lettenmaier, R. J., Stouffer. (2008). “Stationarity is dead: Whither water management.” Science, 319, 573-574.

Misiti, M., Misiti, Y., Oppenheim, G., and Poggi, J., (2010).Wavelet toolbox: For use with MATLAB, The MathWorks, Natic,Mass.,

Moosavi, V., Vafakhah, M., Shirmohammadi, B., and Behnia, N. (2013). “A Wavelet-ANFIS Hybrid Model for Groundwater Level Forecasting for Different Prediction Periods.” *Water Resources Management*, 27(5), 1301–1321.

Nayak, P. C., Venkatesh, B., Krishna, B., and Jain, S. K. (2013). “Rainfall-runoff modeling using conceptual, data driven, and wavelet based computing approach.” *Journal of Hydrology*, Elsevier B.V., 493, 57–67.

Neupauer, R.M., Powell, K.L., Qi, X., Lee, D.H., Villhauer, D.A. (2006). “Characterization of permeability anisotropy using wavelet analysis.” Water Resour. Res. 42, W07419.

Nourani, V., Mogaddam, A.A., Nadiri, A.O. (2008). “An ANN-based model for spatiotemporal groundwater level forecasting.” Hydrol. Process. 22, 5054–5066.

Nourani, V., Alami, M.T., Aminfar, M.H., (2009 a). “A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation.” Eng. Appl. Artif. Intel. 22, 466–472.

Nourani, V., Komasi, M., Mano, A. (2009 b). ”A multivariate ANN-wavelet approach for rainfall-runoff modeling.” Water Resour. Manage. 23 (14), 2877–2894.

Nourani, V., Kisi, O., Komasi, M. (2011). “Two hybrid artificial intelligence approaches for modeling rainfall – runoff process.” Journal of Hydrology, 402, 41-49.

Nourani, V., Baghanam, A. H., Adamowski, J., and Gebremichael, M. (2013). “Using self-organizing maps and wavelet transforms for space–time pre-processing of satellite precipitation and runoff data in neural network based rainfall–runoff modeling.” *Journal of Hydrology*, Elsevier B.V., 476, 228–243.

Pandhiani, S. M., and Shabri, A. B. (2013). “Time series forecasting using wavelet-least squares support vector machines and wavelet regression models for monthly stream flow data.” Open J. of Statistics, 3, 183-194.

Partal, T., and Kucuk, M. (2006 a). “Flow forecasting for a Hawaii stream using rating curves and neural networks.” J. Hydrology, 317, 63-80.

Partal, T., and Kisi, O. (2007). “Wavelet and neuro-fuzzy conjunction model for precipitation forecasting.” J. of Hydrology, 342, 199-212.

Partal, T., and Cigizoglu, H. K. (2008). “Estimation and forecasting of daily suspended sediment data using wavelet-neural networks.” J. of Hydrology, 358, 317-331.

Partal, T. (2009). “Modelling evapotranspiration using discrete wavelet transform & neural networks.” Hydrol. Process. 23 (25), 3545–3555.

Prahlada, R., and Deka, P. C. (2011). “Hybrid wavelet neural network model for improving forecasting accuracy of time series significant wave height.” Int. J. Earth Science and Engg., Vol.4 (5)., Oct. 857-866. (ISSN0974-5904).

Rao, Y. R. Satyaji and Krishna, B. (2009). “Modelling Hydrological Time Series data using Wavelet Neural Network Analysis”, IAHS Publication -333,101-110.

Raghuwanshi, N. S., Singh, R., and Reddy, L. S. (2006). "Runoff and sediment yield modeling using artificial neural networks: Upper Siwane River India." ASCE J. of Hydro. Engineering, 11 (1), 71-79.

Rajaee, T., Mirbagheri, S. A, Nourani, V., Alikhani, A. (2010). “Prediction of daily suspended sediment load using wavelet and neuro-fuzzy combined model.” Int. J. Environ. Sci. Tech., 7 (1), 93-110.

Rajaee, T., Nourani, V., Kermani, M. Z., Kisi, O. (2011). “River suspended sediment load prediction: Application of ANN and wavelet conjunction model”. ASCE J. Hydrologic Engg., 16(8),August,613-627.

Rathinasamy, M., Adamowski, J., and Khosa, R. (2013). “Multiscale streamflow forecasting using a new Bayesian Model Average based ensemble multi-wavelet Volterra nonlinear method.” Journal of Hydrology, Elsevier B.V., 507, 186–200.

Rioul, O, and Vetterli, M. (1991). “Wavelets and signal Processing.” IEEE SP Magazine, 14-38.

Salvatore, C. P., and Adamowski J. (2012). “Forecasting urban water demand via wavelet denoising and neural network models. Case study: City of Syracuse, Italy.” Water Resour. Manage., 26, 3539-3558.

Sang, Y.F., Wang, D. (2008). ”Wavelets selection method in hydrologic series wavelet analysis.” J. Hydraul. Eng. 39 (3), 295–300 (in Chinese with English abstract).

Sang, Y.-F. (2012). “A Practical Guide to Discrete Wavelet Decomposition of Hydrologic Time Series.” *Water Resources Management*, 26(11), 3345–3365.

Sharma, J.N. (2005) “Fluival processs and morphology of the Brahmaputra river in Assam, India”. Geomorphology, 70, 226-256.

Shirmohammadi, B., Moradi, H., Moosavi, V., Semiromi, M. T., and Zeinali, A. (2013). “Forecasting of meteorological drought using Wavelet-ANFIS hybrid model for different time steps (case study: southeastern part of east Azerbaijan province, Iran).” *Natural Hazards*, 69(1), 389–402.

Shiri, J., and Kisi, O. (2010). “Short term and long term stream flow forecasting using a wavelet and neuro-fuzzy conjunction model.” Journal of Hydrology, 394, 486-493.

Solaimani, K. (2009). “Rainfall-runoff Prediction Based on Artificial Neural Network; A Case Study: Jarahi Watershed.” American-Eurasian J. Agric. & Environ. Sci., 5 (6), 856-865

Sudheer, K. P., Gosain, A. K., Ramsastri, K. S. (2002). "A data-driven algorithm for constructing artificial neural network rainfall-runoff models." Hydrological Processes, 16, 1325-1330.

Surinder Deswal, and Mahesh Pal, (2008). “Artificial Neural Network based Modeling of Evaporation Losses in Reservoirs”, World Academy of Science, Engineering and Technology, 39, 279-283.

Tayfur, G., and Singh, V. P. (2006). “ANN and fuzzy logic models for simulating event based rainfall-runoff.” J. of Hydrologic Engg., 132 (12), 1321-1330.

Torrence, C., Compo, G.P. (1998). “A practical guide to wavelet analysis.” Bull. Of American Meteorological Soc. 79 (1), 61–78.

Venkata Ramana, R., Krishna, B., Kumar, S. R., and Pandey, N. G. (2013). "Monthly Rainfall Prediction Using Wavelet Neural Network Analysis." *Water Resources Management*, 27(10), 3697–3711.

Wang, W., Yuan, P., Ding, J. (2000). "Wavelet analysis and its applications to the stochastic simulation of daily discharge process." J. of Hydraulics, 11;43-48.

Wang, W., Ding, J., Xiang, H. (2002). "The multi-time scale analysis of hydrological time series with wavelet transform." J. of Sichuan university, 35(4),14-17.

Wensheng, W., and Jing, D. (2003). "Wavelet Network Model and Its Application to the Prediction of Hydrology." Nature and Science 1(1), 67-71.

Wang, W., Hu, S., and Li, Y. (2011). "Wavelet Transform Method for Synthetic Generation of Daily Streamflow." *Water Resources Management*, 25(1), 41–57.

Wang, Y., Wang, H., Lei, X., Jiang, Y., and Song, X. (2011). "Flood simulation using parallel genetic algorithm integrated wavelet neural networks." Neurocomputing, 74, 2734-2744.

Wu, J. S., Han, J., Annambhotla, S., Bryant, S. (2005). "Artificial Neural Networks for Forecasting Watershed Runoff and Stream Flows." J. of Hydrologic Engg., ASCE, 10 (3), 216-222.

Zadeh, M. R., Seifollah, A., Khalili, D., Singh, V. P. (2010). "Daily Outflow Prediction by Multi Layer Perceptron with Logistic Sigmoid and Tangent Sigmoid Activation Functions." Water Resour Manage, (2010) 24:2673–2688.

Zhou H. C, Peng, Y, Liang, G. H. (2008). "The research of monthly discharge predictor-corrector model based on wavelet decomposition." Water Resources Management, 22:217-227.

LIST OF PUBLICATIONS BASED ON Ph.D. RESEARCH WORK

Papers in Refereed International Journals:

1. **Khandekar Sachin Dadu** and Paresh Chandra Deka (2012).
   "Wavelet-neural network conjunction model in flow forecasting of sub-Himalayan river Brahmaputra." Int. J. of Civil Engg. and Technology, 3 (2), 415-425.
2. **Khandekar Sachin Dadu** and Paresh Chandra Deka (2013).
   Multistep lead time forecasting of hydrologic time series using Daubechies wavelet-neural network hybrid model, International Journal of Scientific and Engineering Research, Volume 4, Issue 10, pp 115-124.

Papers in International Conferences:

1. **Khandekar Sachin Dadu** and Paresh Chandra Deka (2012).
   "Applications of wavelet transform technique in Hydrology – a brief review." International conference on Environmentally Sustainable Urban Ecosystems "ENSURE2012", Conference hosted by IIT Guwahati, India.

Papers in National Conferences:

1. **Khandekar Sachin Dadu** and Paresh Chandra Deka (2013).
   "Wavelet-neural network conjunction model for Brahmaputra River flow forecasting." National Conference on Sustainable Water Resources Development and Management 'SWARDAM 2013', Conference hosted by Govt. College of Engg. Aurangabad, Maharashtra, India.

BIO-DATA

Name	KHANDEKAR SACHIN DADU
Designation	Research Scholar
Date of birth	19th July 1977
Email	khandekarsd@gmail.com
Contact No.	09011092396
Permanent Address	“Grand View 7”, H -202, Near Podar International School, Ambegaon (Bk.), Pune - 411046
Educational Qualification	B.E. Civil M.Tech. Civil (Hydraulics & Water Resources Engg.)
Current Work Place	Asst. Prof. of Civil Engg., STES’s Sinhgad Institute of Technology & Science, Narhe, Pune – 411041.