STREAMFLOW FORECASTING USING WAVELET COUPLED SOFT COMPUTING TECHNIQUES AND FUZZY LOGIC-BASED APPROACH FOR STREAM WATER QUALITY-QUANTITY ASSESSMENT

Thesis

Submitted in partial fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

by

SHRUTI KAMBALIMATH S

Register No. -- 177058AM009

DEPARTMENT OF WATER RESOURCES AND OCEAN ENGINEERING

NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA

SURATHKAL, MANGALORE-575025

FEBRUARY 2023

<u>**D E C L A R A T I O N**</u>

I hereby *declare* that the Research Thesis entitled **“Streamflow forecasting using wavelet coupled soft computing techniques and fuzzy logic-based approach for stream water quality-quantity assessment”** which is being submitted to the **National Institute of Technology Karnataka, Surathkal** in partial fulfilment of the requirements for the award of the Degree of **Doctor of Philosophy** in the Department of Water Resources and Ocean Engineering is a *bonafide report of the research work carried out by me*. The material contained in this Thesis has not been submitted to any University or Institution for the award of any degree.

SHRUTI KAMBALIMATH S
177058AM009
Department of Water
Resources and Ocean Engineering
NITK, Surathkal

Place: NITK, Surathkal

Date: 24/02/ 2023

CERTIFICATE

This is to *certify* that the Research Thesis entitled **"Streamflow forecasting using wavelet coupled soft computing techniques and fuzzy logic-based approach for stream water quality-quantity assessment"** submitted by **SHRUTI KAMBALIMATH S, (Register Number: 177058AM009)** as the record of the research work carried out by him/her is *accepted as the Research Thesis submission* in partial fulfillment of the requirements for the award of the degree of **Doctor of Philosophy**.

Research Guide: 24/02/2023

Prof. Lakshman Nandagiri

Professor

Department of Water Resources and Ocean Engineering,

NITK, Surathkal.

(Name and Signature with Date and Seal)

Dept. of Water Resources & Ocean Engineering NITK Surathkal

Chairman – DRPC

(Signature with Date and Seal)

Chairman (DRPC)

Dept. of Water Resources & Ocean Engineering

ACKNOWLEDGEMENT

It gives me immense pleasure and innate satisfaction to express my deep sense of gratitude and indebtedness to my revered Prof. Paresh Chandra Deka to whom I owe what I have learnt in the field of Hydrology and Water Resources Engineering. His expert guidance, training and encouragement throughout this journey have made this work possible. He stood as the biggest moral support when I had to go through the toughest times of my personal life during my PhD journey. I thank him wholeheartedly for showing such kindness and right guidance at crucial times. I feel very unfortunate for the loss of our beloved research supervisor but he sure will as always bless his students and guide them even without being present physically.

I sincerely express my heartfelt thanks to my research supervisor Prof. Lakshman Nandagiri, who accepted me as his student during crucial time of this journey. His moral support, guidance, interactions, discussion and precious suggestions have greatly helped me to complete this work. I express my wholehearted gratitude for his timely responses and expert guidance.

I would like to convey my sincere thanks to my Research Progress Assessment committee (RPAC) members, Dr. Subrahmanya K., Department of Water Resources and Ocean Engineering and Dr. Arun Kumar Thalla, Department of Civil Engineering, NITK for providing valuable suggestions, comments and encouragement at various stages of this research work.

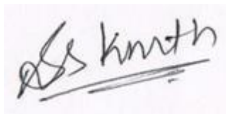
My express my sincere thanks to Prof. B M Dodamani, HOD, Department of Water resources and Ocean engineering. I also thank all the teaching faculty and non-teaching staff of the Department of Water resources and Ocean engineering for their support throughout this journey. I sincerely thank the Director of the Institute (NITK), for the financial support provided and for extending the facilities for the research.

I take this opportunity to specially thank my loving family. I am very grateful to my husband Mr. Shivalingaswami Halagalimath, who stood with me and supported me during all the critical times of this journey. Words cannot express how grateful I am to my beloved father Mr. Shivaprakash Kambalimath, who dreamt of my PhD and whom I lost during this journey. His love and blessings are always with me with which I am able to complete this work. I thank my mother Mrs. Sudha Kambalimath and my sister Mrs. Keerti Kambalimath for all the support and all the sacrifices that they have made for me. A special thanks to my little daughter Shaarvari, who is the sweetest gift from almighty. I consider myself to be very lucky and blessed to have such a loving and supportive family. I also thank my father-in-law Mr. Shantaveerayya Halagalimath,my mother-in-law Late. Mrs. Shashikala Halagalimath and my mother-in-law Mrs. Uma Halagalimath.

I thank my friends and colleagues for all their support during the journey. I thank my dearest friend Mr. Nataraja M. for his kind support throughout this journey. I specially thank Mr. Gunjan Kumar Nayak for his support and valuable suggestions for the coding part of the research work, Mr. Bhuvanamitra, Mr. Mallikarjun Bandiwad, Mr. Parthasarathy and Mrs. Ashwitha for their moral support throughout this journey. I thank my neighbour and senior Dr. Basavaraj Nainegali and his wife Mrs. Sapna, who have been a helping hand from the beginning of this journey till the end.

I sincerely thank the Water Resources Department, Government of Karnataka and Chief Engineers Office, Malaprabha Zone, Dharwad for providing valuable data resources for my research work.

Last but not the least, I thank the Almighty for giving me enough strength to fight all the struggles and pursue my dream. I thank my Durgaparameshwari amma for bringing the positive energy within me.



SHRUTI KAMBALIMATH S

DEDICATED TO MY PARENTS, MY HUSBAND AND MY RESEARCH GUIDE THE LATE. DR. PARESH CHANDRA DEKA

ABSTRACT

Hydrologic time series is a collection of timely recorded variables such as streamflow, temperature, evaporation, etc. over a period of time. Forecasting of such time series necessarily aid future predictions based on past records as well as filling of missing data or extension of available data. Accurate and timely forecasting of hydrologic time series can be a great aid for various applications in water resources planning and management.

During the last few decades, several types of stochastic models have been proposed as well as developed for modeling hydrological time series and generating synthetic stream flows. Some of such stochastic models are autoregressive (AR), Moving Average (MA), Autoregressive Moving Average (ARMA), and Autoregressive Integrated Moving Average (ARIMA).

In contrast to the analytical models, soft computing methods learn from past records and require limited input parameters. These techniques are very useful in cases where there are limitations in terms of data availability. The collection of techniques under Soft Computing renders low-cost solutions to imprecisely formulated problems and attempt to mimic the behavior and learning ability of human beings into computers. One such soft computing technique is “Fuzzy Logic”. We have developed three soft computing models to forecast daily streamflow time series for different lead times for Malaprabha sub-basin in Karnataka state of India. The performance of Support Vector Machine (SVM), Adaptive Neuro-fuzzy Inference System (ANFIS), and Fuzzy models to forecast daily streamflow is tested for 1-day, 3-days and 5-days ahead forecasts. The results indicate that the performance of the models significantly decreases with an increase in lead times. The models show high R^2values for 1-day ahead streamflow forecasts, whereas it is low for 3 and 5-days lead time.

It is necessary to provide a powerful tool to reduce the noise in the data so that accuracy of the model is increased. Wavelet transformer is one such powerful tool used to decompose the data set into different scales. The wavelet method effectively decomposes the original time series in to sub-series at different resolution levels there by facilitating denoising of the data. In this research work, discrete wavelet transform is coupled with the fuzzy logic method to improve the accuracy of the forecast. The performance of all the three models significantly increased when the wavelet is coupled, especially for longer lead times such as 5 days. The Wavelet coupled fuzzy (WT-fuzzy) model outperformed Wavelet coupled ANFIS (WT-ANFIS) and Wavelet coupled SVM (WT-SVM) models. However, WT-ANFIS performed better than WT-SVM.Longer lead time forecasts find applications in flood forecasting and evacuation programs. This research aims at improving the efficiency of forecasting models especially for longer lead times such as 3 days and 5 days which are crucial times for undertaking quick flood evacuation measures.

The second phase of this research is stream water quality-quantity modeling. Water quality and quantity are the two aspects that are interrelated and hence should be studied together within an integrated framework. In today's world, demand for water essentially takes into account both quality and quantity aspects for various uses of water. Having a sufficient accessible quantity of water becomes meaningful only if this quantity of water is acceptable in terms of its quality. This study aims at studying the role of the quantity of water in determining its quality along with the other quality parameters. The Water Quality Index (WQI) is an efficient tool which can describe the status of water by translating a large amount of data in to a single value. The results in this study indicate that streamflow can be considered as one of the inputs to determine the WQI.

Keywords: Soft computing, Fuzzy Logic, Support Vector Machine, Adaptive Neuro-fuzzy Inference System, Wavelet Transform, Stream water quality-quantity, Water Quality Index.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOTATIONS AND ABBREVIATIONS

Abbreviation/Notation	**Definition**
AR	Autoregressive
MA	Moving Average
ARMA	Autoregressive Moving Average
ARIMA	Autoregressive Integrated Moving Average
ARMAX	ARMA with exogenous input
ARIMAX	ARIMA with exogenous input
SVM	Support Vector Machine
ANFIS	Adaptive Neuro-Fuzzy Inference system
FL	Fuzzy Logic
SVR	Support Vector Regression
AI	Artificial Intelligence
ML	Machine Learning
GA	Genetic Programming
GP	Genetic Algorithm
ANN	Artificial Neural Network
NN	Neural Network
WT-ANFIS	Wavelet coupled ANFIS
WT-FUZZY	Wavelet coupled FL
WT-SVM	Wavelet coupled SVM
WQI	Water Quality Index
SWAT	Soil and Water Assessment Tool
UBCWM	UBC Watershed Model
FES	Fuzzy Expert System
MF	Membership Function
WT	Wavelet Transform
PSO	Particle Swarm Optimization
WD	Wavelet Decomposition
W-ANN	Wavelet coupled ANN
W-ANFIS	Wavelet coupled ANFIS
W-GP	Wavelet coupled GP
FWQI	Fuzzy Water Quality Index
MLR	Multiple Linear Regression

WNN	Wavelet Neural Network
NW	Neuro-wavelet
FFNN	Feed Forward Neural Network
WARM	Wavelet Autoregressive Model
DRNN	Dynamic Recurrent Neural Network
TRBP	Temporal Recurrent Back Propagation
DWT	Discrete Wavelet Transform
CWT	Continuous Wavelet Transform
FNN	Fuzzy Neural Network
SWMM	Storm Water Management Model
R-R	Rainfall-Runoff
NFS	Neuro Fuzzy System
NF	Neuro-Fuzzy
WNF	Wavelet Neuro-Fuzzy
WFL	Wavelet Fuzzy Logic
WANFIS	Wavelet-ANFIS
HEC-HMS	Hydrologic Engineering Centre's Hydrologic Modeling System
GANN	Genetic Algorithm Neural Network
FIS	Fuzzy Inference System
BOD	Biological Oxygen Demand
COD	Chemical Oxygen Demand
DO	Dissolved Oxygen
Tco/TC	Total Coliforms
Temp	Temperature
WRIS	Water Resources Information System
TDS	Total Dissolved Solids
Ec	Electrical Conductivity
Cl	Chlorine
Na	Sodium
K	Potassium
SO_4	Sulphate
NO_3	Nitrate
$CaCO_3$	Calcium Carbonate
NH_3	Ammonia

NH_4	Ammonium
CO_2	Carbon Dioxide
Mg	Magnesium
Ca	Calcium
SAR	Sodium Adsorption Ratio
Db	Daubechies
Coif	Coiflets
Syms	Symlets
RBF	Radial Basis Function
SMO	Sequential Minimization Optimization
R^2	Correlation Coefficient
RMSE	Root Means Square Error
NSE	Nash-Sutcliffe Efficiency
FC	Fecal Coliform
MPN	Most Probable Number
FAO	Food and Agricultural Organization
BIS	Bureau of Indian Standards
ACF	Autocorrelation Function
PACF	Partial Autocorrelation Function
mg/L	Milligram per litre
m^3/s	Metre cube per second
µmho/cm	Micro mho per centimetre
ψ	Psi
ω	Omega
ξ	Xi
ε	Epsilon
ϕ	Phi
α	Alpha
$\in$	Belongs to
$\sum$	Summation
Σ	Sigma

CHAPTER 1

INTRODUCTION

1.1 Background

Predictions of historical data records play a very important role in systematic planning and assessing of the design criteria of water resource structures. The recognition of appropriate generation models for potential streamflows is an essential prerequisite for efficient and reliable planning and management of water resources. Especially, problems like filling in missing data as well as the expansion of existing data records can be accomplished by using the methods like synthetic streamflow generation models. These types of simulation methods facilitate to forecast possible replicates of potential streamflows for the design hydrologist (Keskin et al. 2004).

Precise and acceptable prediction of hydrological time series pertaining to temperature, rainfall, evaporation, streamflow, etc. has driven great attention in the field of water resources engineering. It can be one of the important aids to planners and managers for planning and optimal usage of available water resources. Hydrologic time series forecasting investigations in the past decades has shown the development of a large number of methods and models and a variety of approaches have been reported in order to improve the accuracy of forecasting. The models developed so far can be categorized into physical methods, statistical methods, and artificial intelligent approaches. However, none of the single forecasting methods can be considered superior for any hydrologic time series since the hydrological systems are naturally affected by several factors related to evapotranspiration, climate, soil infiltration rates and land cover. Consequently, hydrologic time series exhibit stochastic components and nonlinear characteristics which operate at multiple spatio-temporal scales. Accordingly, each one of the categories of methods has different advantages and short comes (Machiwal and Jha 2012).

One of the fundamental topics in stochastic hydrology is the hydrologic time series analysis, which intends at enlightening complex hydrologic processes. It is a very essential task and in exercise it forms the foundation for hydrologic model simulation and forecasting, management of water resources, as well as several other waters related analysis and designs. Hydrologic time series analysis is however, also a difficult assignment for the reason that it is influenced by several unfavourable conditions. Adequate understanding of the processes involved in the hydrologic systems has not been achieved at present, and usually the deterministic components present in the observed historical hydrologic data are unknown. Furthermore, noise is an unavoidable part of the observed hydrological data that causes a variety of complications in the analysis of hydrologic series, such as hydrologic forecasting, parameter estimation and period identification. As a result, exact results of hydrologic time series analysis cannot be achieved effortlessly (Machiwal and Jha 2012).

The advancement of computer systems and growth in software and hardware has shown the way to the rapid emergence and growth in the domain of computational or artificial intelligence. The growth of computational intelligence has fetched an innovative modification in the advancement of novel non-conventional methods of data simulation and processing. Assimilation of intelligence by imitating the human intelligence, reasoning and behavior into the computing systems improves its competency to examine the information which is subjected to a dynamically changing environment (Chadwani et al. 2015).

The term “Soft computing” includes a group of computational techniques motivated by inbuilt vagueness, intuition, wisdom and consciousness of human beings and the uncertainty involved in real-life. Unlike the traditional computing methods which depend on accurate solutions, soft-computing techniques aim at utilizing the given insignificant and uncertain nature of the problem and the tolerance of imprecision to yield a quick and approximate solution to a given problem. Soft Computing domain is a multi-disciplinary field which utilizes various methods of probabilistic, statistical and optimization techniques which balance each other to develop distinctive computational strategies namely, Neural Networks, Evolutionary Computation, Fuzzy Systems, Machine Learning and Probabilistic Reasoning. Among several sub-sets of the soft computing field, Fuzzy Logic, Genetic Algorithms and Neural Networks are the foremost players and are frequently used for problems associated to real life applications (Chadwani et al. 2015).

River streams and lakes are one of the major water resources used for human consumption and domestic use (Kukrer and Mutlu 2019). Contamination of such water resources lead to significant problems related to public health and environment. Natural processes such as rain, abrasion, soil erosion, etc. and anthropogenic activities such as urban industrial, agricultural activities, etc. pose a serious pressure on surface water quality (Wu et al. 2018; Zhao et al. 2012, Kukrer and Mutlu 2019). In addition, climate change also affects the aquatic systems (Wu et al. 2015; Kukrer and Mutlu 2019).

Water quality for a given river is considered to be the result of numerous inter-related variables with variations at local and temporal scales influenced by the flow rate of water throughout the year (Dunca 2018). Water pollution results in the reduction of the available quantity of fresh water for both population and ecosystems, thereby contributing to the "Global water crisis" (Ganoulis 2009).

Water quality and quantity are the two aspects which are inter-related and hence should be studied together within an integrated framework. In today's world, demand for water essentially takes into account both quality and quantity aspects for various uses of water (Ganoulis 2009). Having sufficient accessible quantity of water becomes meaningful only if this quantity of water is acceptable in terms of its quality (Ahmed et al. 2015). This study aims at studying the role of quantity of water in determining its quality along with the other quality parameters.

Changes in the flow pattern and flow rate of water in rivers lead to changes in the dilution rate of nutrients, thereby altering the overall quality of water in streams. Hence it turns out to be important to study the quality of water for different flow rates throughout the year. This study makes an attempt to examine the variation of water quality which is expressed in terms of a water quality index (WQI) with the quantity of water for a particular time period.

Many researchers have examined and reported the significance of water quality indices (WQI) which provides an indication for the evaluation of water quality as first proposed by Horton (1965). Later on, several researchers have carried out investigations on water quality evaluation using WQIs all around the world in several water bodies (Kangabam et al. 2017). Nevertheless, literature was not found which revealed the detailed and scientific study carried out related to WQI in Malaprabha sub-basin, which is a major river in North Karnataka.

1.2 Modeling Hydrologic Time Series

Modeling and forecasting of time series have gained vast attention from many researchers in recent times. The reason for such a tremendous development in the modeling field is that the forecasts of a hydrological variable, which is usually observed as discrete or continuous variable for a particular time period, are critically important for planning, designing and managing the various water related activities.

The conventional methods of time series modelling and forecasting include the well-known “Box-Jenkins” techniques of AR, ARMA, ARIMA, ARMAX, etc (Jain and Kumar 2007). In some of the cases of hydrological modelling of time series, a simple basic stochastic model performed very well compared to a complicated deterministic model (Hipel 1985; Lohani et al. 2012). The AI methods, especially the soft computing techniques combine methodologies from different sources, model the systems with human-like expertise, adapt and learn by themselves in a changing environment.

Streamflow is one of the important hydrologic variables, which is measured at a given location in a river, and expressed in terms of m^3/s. The future predictions of streamflow values at a given location in a basin is very important for managing and designing of many water resources related projects (Jain and Kumar 2007). Although these soft computing techniques do not present any results based on the physical aspects of hydrological processes, they are still very suitable for stream flow flood forecasting where the primary focus is to accurately predict flow at a given location in a watershed (Nayak et al. 2005; Lohani et al. 2012).

1.3 Soft computing techniques in hydrologic time series forecasting

During the last few decades, several types of stochastic models have been developed and proposed for modeling hydrological time series and generating synthetic stream flows. Some of such stochastic models are autoregressive (AR), Moving Average (MA), Autoregressive Moving Average (ARMA), and Autoregressive Integrated Moving Average (ARIMA).

The umbrella of Soft Computing techniques is widening its scope with every passing year due to increasing demand for time saving and fault tolerant computational tools. In contrast to analytical methods, Soft Computing methodologies mimic consciousness and cognition in several important aspects: they learn from experience; they can universalize into domains where direct experience is absent; and, through parallel computer architectures that simulate biological processes, they can perform the mapping from inputs to the outputs faster than inherently serial analytical representations (Chaturvedi 2008). The collection of techniques under Soft Computing render low-cost solutions to imprecisely formulated problems and attempt to inculcate the behavior and learning ability of human beings into computers. The following sub-sections deal with an introduction to the forerunner techniques of Soft Computing viz., Artificial Neural Networks (ANN), Support Vector Machines (SVM), Genetic Programming (GP) and Fuzzy Logic (FL).

1.3.1 Artificial Neural Networks (ANNs)

The Neural Networks (NN) or popularly known as the Artificial Neural Networks (ANNs) corresponds to the architectures and learning algorithms motivated by the structure, operation and functioning of the human being's brain. Even though they signify a much simplified and basic version of the human brain, these neural network models have given a new direction to resolve many problems occurring in real life processes. Haykin (2009) explained neural networks as an extraordinarily parallel spread processor which is combined of simple smaller units of processing that exhibit a natural tendency for accumulating observed information and reproducing it for the intended use. These NNs represent a data learning prototype consisting of processing units known as "artificial neurons" or "neurons" that are organized in different layers. Independently, these neurons execute minor functions whereas jointly, they form a network, which are efficient of solving much more complicated real-life problems and issues (Flood and Kartam 1994).

The structural design of ANNs comprise of three vital components and functions viz., the weighted relations linking the neurons internally, the associated learning algorithm which can update these weights at each iteration and the activation function that acts on the total weighted sum of the input data signal taken by the neuron. These ANNs are fit well for those problems whose solutions involve information that is complicated to identify but have sufficient observations or data (Zhang et al.1998). ANN's capability to present model free from complications and boundary conditions, the parallel processing of data associated with noise and adaptableness to altering conditions of the given problem, provides with an edge over traditional data processing methods. The multi-disciplinary applications of ANN are attributed to its ability of deriving complex, non-linear and unknown relationships among independent and dependent variables through a learning process, thereby working as a universal function approximator and therefore has been a field of interest for predicting the behavior of engineering and natural systems (Chadwani et al. 2015).

1.3.2 Support Vector Machines (SVM)

Support Vector Machines (SVM) and their extension, Support Vector Regression (SVR), constitute another soft computing tool for non-linear regression and have been used in hydrological research for nearly a decade (Hsieh 2009). The procedure is similar to ANNs, where inputs are mapped non-linearly to a hidden space and then to an output space, and the model attempts to minimize an error function through training. In SVMs, these spaces are transformed into a much higher dimension using a mathematical formula known as a kernel function. In this higher-dimensional space the mapping functions become linear and the dimensionality of the error function is similarly reduced (Bourdin et al. 2012).

Many of the limitations of ANNs are improved upon by SVMs, and they have been shown to outperform ANNs in many hydrological applications. For example, Behzad et al. (2009) reported that SVMs are able to generalize better than ANNs, though there is still some danger of under or over-fitting to the training data (Han et al. 2007) (true to some extent of virtually any model). The SVMs are also able to learn from a much smaller training set than ANNs, and the global minimum of the linear optimization is easily obtainable, whereas there is a risk of becoming trapped in a local minimum of the non-linear ANN objective function (Behzad et al. 2009; Bourdin et al. 2012).

Despite being a relatively new approach to hydrological modelling, SVMs and SVRs have been applied to many of the same problems as ANNs, including rainfall-runoff modelling for water resources planning and flood forecasting at various lead-times (Asefa et al. 2006; Han et al. 2007; Behzad et al. 2009; Rasouli et al. 2012; Bourdin et al. 2012), hydraulic modelling (Liong and Sivapragasam 2002;Bourdin et al. 2012) and downscaling of GCM output (Tripathi et al. 2006;Bourdin et al. 2012).

1.3.3 Genetic Programming (GP)

Genetic Programming (GP) (Koza 1992) is another soft computing approach to non-linear modelling and is based on Darwin's theory of evolution by natural selection. GP is a recent variant on evolutionary methods such as genetic algorithms (GA), which have been used to optimize ANN architecture and weights (Dawson and Wilby 2001), SVM parameters (Tripathi et al. 2006), and the parameters of physically oriented hydrological models such as the UBCWM and SWAT (Lan 2001; Zhang et al. 2009). In GP, a random initial population of equations relating predictors to predictands undergoes evolution over many generations through the sharing and mutation of genetic material (predictors and mathematical operators). The "best" individuals from each generation are selected based on some measure of fitness and go on to produce further generations. Only important variables are retained through the evolutionary process. The input variables and mathematical or logical functions that can be combined into equations are chosen by the model developer based on his or her understanding of the processes being modeled (Bourdin et al. 2012).

1.3.4 Fuzzy Logic (FL)

Modeling processes are always expected to have precision which is not the case of real-world problems. The problems that the real-world deals are often involved with imprecision, vagueness and lack of clarity. Such problems can be handled by Fuzzy Expert Systems (FES) that is based on the technique called fuzzy logic (Zadeh 1965). This is achieved through a systematic approach of simulation of the process of human thinking. So far, the computer systems are programmed to utilize the approximate or crisp reasoning which are referred to as black box and white box (Boolean) models (Shu and Burn 2004; Liao 2005). The foremost step in developing a fuzzy inference system is to select the input as well as output variables and each of the variables is defined by a fuzzy set, which are expressed in terms of linguistic variables, with associated membership functions (MF) that are associated with quantitative values. For instance, rainfall variable can be classified as high, low or average and at a time, a given value may be simultaneously holding partial membership in one or more of the above-mentioned categories. The next step is the formations of rules that can associate the membership functions to the outputs and are well defined by simple verbal or linguistic statements with conditions based on the expert knowledge. The last step is an iterative one which evaluates and tunes relations of observations to the rules that are formed in the previous step. Finally, the formed fuzzy rules are well combined with the help of an inference system engine, and here the defuzzification method is applied to convert the fuzzy or linguistic model output variables into the single crisp values.

Very similar to Genetic Programming (GP), FES has the benefit of transparency, as the knowledge of expert system is applied to describe the membership function rules (Chau et al. 2005). The results of such an FES are very much independent of the amount of data on hand for the purpose of training (Mahabir et al. 2003). The FES systems have been used with acceptable results in applications like flood frequency analysis studies (Shu and Burn 2004), optimal reservoir planning and operations and river flow forecasting (Russell and Campbell 1996), also in seasonal water supply forecasting (Mahabir et al. 2003) and modeling of the single components of the hydrological cycle, like infiltration and surface runoff modeling (Bárdossy 1996). Hopeful results have also been achieved through the combinations of fuzzy logic and ANNs for short-term rainfall-runoff modeling studies (Chang and Chen 2001).

1.4 Wavelet Transform

Wavelet transforms (WT) have recently begun to be explored as a tool for the analysis, de-noising and compression of signals (e.g., time series) and images. WT separates a signal into shifted and scaled version of the original (or mother) wavelet. In other words, by a WT, a signal decomposes into multiple levels of details, sub-signals, which provide an interpretation of the time series structure and history in both the time and frequency domains using a few coefficients (Rajaee et al. 2010). Wavelets are wave-like mathematical functions with amplitude that begins at zero, increases, and then decreases back to zero. Unlike the sine waves, they generally tend to be irregular and asymmetric (Ozger 2010). WT allows the use of long-time intervals for low frequency signals and shorter intervals for high frequency signals and is able to reveal some statistical features of time series like trend and shift that other signal analysis techniques such as Fourier transform might miss. Another advantage of WT is the flexible choice of mother wavelet according to the characteristics of the investigated time series (Adamowski and Sun 2010). Mallat (1998) provides detailed information about wavelet functions.

1.4.1 Wavelet - Fuzzy hybrid approach in hydrological time series

Artificial Intelligence methods like Fuzzy Logic, ANNs, GP have been applied in many studies for the analyses of stationary and non-stationary time series. However, it is noted that soft computing technique single-handedly is not adequate to resolve the complicated problems such as missing data, limited data, since such data are noisier. Therefore, it is essential to provide a dominant tool in order to decrease the noise in the data so that precision of the model is increased. Wavelet transformer is one such powerful tool which can be used to decompose the data set into different scale. In this investigation, discrete wavelet transformer is coupled with fuzzy logic method to improve the accuracy of the forecast. Thus, combining wavelet technique with fuzzy logic may increase the forecasting accuracy (Surendra and Deka 2015).

Wavelet analysis is the breaking up of a signal into shifted and scaled versions of the original (or mother) wavelet. In wavelet analysis, the use of a fully scalable modulated window solves the signal-cutting problem. The window is shifted along the signal and for every position the spectrum is calculated. Then this process is repeated many times with a slightly shorter (or longer) window for every new cycle. In the end, the result will be a collection of time-frequency representations of the signal, all with different resolutions. Because of this collection of representations, we can speak of a multi-resolution analysis. By decomposing a time series into time-frequency space, one is able to determine both the dominant modes of variability and how those modes vary in time. Wavelets have proven to be a powerful tool for the analysis and synthesis of data from long memory processes. Wavelets are strongly connected to such processes in that the same shapes repeat at different orders of magnitude. The ability of the wavelets to simultaneously localize a process in time and scale domain results in representing many dense matrices in a sparse form (Budu 2013).

Fuzzy time series prediction is a prudent avenue in the areas where information is inexplicit, unclear and approximate. Also, fuzzy time series can tackle circumstances which do not provide the study and analysis of trends nor the visualisation of patterns in time series. Profound research work has been accomplished on forecasting problems using this concept. It is clear from over two decades of research on the relative accuracy of various extrapolative methods that simple methods will in most forecasting situations, and for most data types, produce the most accurate forecasts.

In order to overcome the limitations of pure fuzzy modeling, a hybrid wavelet-fuzzy modeling approach has been considered for this study which can improve the accuracy of hydrologic time series forecasting.

1.5 Stream Water Quality-Quantity Modeling

Water is a very essential natural resource for human beings as well as the health of its environment. Therefore, its quality is very important aspect. Assessment of surface water quality is a very sensitive topic which is also a great environmental concern worldwide. Surface water pollution by chemical, physical, microbial and biological contaminants can cause epidemic problems, at times all over the world (Singh et al. 2005). Fish survival / growth and other biodiversity, conservation activities, recreational activities like swimming and boating, industrial / municipal water supply, agricultural uses such as irrigation and livestock watering, waste disposal and all other water uses are affected by the physical, chemical, microbial and biological conditions that exist in the water courses and also in subsurface aquifers.

The surface water systems are naturally open to the atmosphere, such as lakes, rivers, estuaries, reservoirs and coastal waters. A natural process such as changes in erosion, precipitation, weathering of crustal material as well as any anthropogenic influences such as urban, industrial and agricultural activities, increasing rate of consumption of water resources, degrade the quality and quantity of surface water and make it unsuitable for domestic uses. Industrial waste water, runoff over the agricultural lands and municipal sewage disposal are the most vulnerable for water pollution (Singh et al. 2005). The concentration of biological available nutrients in excess and concentration of toxic chemicals leads to diverse problems such as toxic algal blooms, loss of oxygen in water, fish kill loss of biodiversity and loss of aquatic plants and coral reefs (Vousta et al. 2001).

Water quality modeling is challenging due to some limitations including sufficient representative site selection and sample gaps, and lack of calibration, errors in data reporting and data length. The uncertainty in water quality modeling commonly comes from numerous sources of errors: measurements of input and response uncertainty, parametric uncertainty, and structural error due to the incapability of a specified model structure to reproduce the physical mechanisms. Observed or measured data are typically rare and not sufficient to indicate or predict a complete picture of the real scenario in the large and complex waterbody (Ejigu 2021)

1.6 Scope of the present study

In the present study, the applications of forecasting are differentiated in to long-term (5-days ahead) and short-term streamflow forecasting (1-day and 3-days ahead). As per the literature survey, the accuracy of forecasting models degrades for longer lead times. The purpose of testing the performance of all the three models is to evaluate the best suitable forecasting model for longer lead times. The long-term forecasting finds application in flood evacuation, water resources system planning and management, flood hazard mapping etc.

Hybrid models combine the data-driven models with a data pre-processing technique. Widely-used data pre-processing methods include PSO, GA and WD (Magar and Jothiprakash 2011). The wavelet method can decompose data into a series of signals which help abstract useful information from data and hence reduce uncertainty of forecasting results. That is why hydrological forecasting by combining wavelet decomposition and a data-driven model has received much more attention in recent years (Zhang et al. 2018).

A number of combinations of wavelet hybrid models like W-ANN, W-ANFIS, W-GP, etc., for streamflow forecasting has been developed by the researchers. But the combination of Wavelet-Fuzzy is considerably new and hence it is a major focus of the study. This study demonstrates the improvement in the results of forecasting model when coupled with a data pre-processing technique like wavelet transform.

The second major part of the study is stream water quality-quantity modeling. There are very few literatures which connect stream water quality and quantity. This study demonstrates this connection by using a fuzzy water quality model (FWQI).

1.7 Organization of the thesis

This thesis is organized in to six chapters as follows:

Chapter 1 provides an overview of hydrologic time series analysis and forecasting. A brief description is given regarding the various methods of hydrologic forecasting with particular emphasis on streamflow forecasting and its methods. It also includes the scope of the present study.

Chapter 2 presents a detailed review of literature related to hydrologic forecasting, soft computing techniques and hybrid models used for streamflow forecasting and stream water quality modeling. Based on the literature review, research gaps are identified and the study objectives are formulated accordingly.

Chapter 3 provides a description of the study area and the justification for selecting the study area. A description of the data collected for analysis is also provided.

Chapter 4 discusses in detail, the methods like ANFIS, SVM, Fuzzy Logic and wavelet decomposition used in the study for streamflow forecasting. The detailed mathematics underlying the above-mentioned methods and detailed flowchart of methodology are also incorporated in this chapter.

Chapter 5 presents the results of streamflow forecasting and stream water quality-quantity modeling. The performance of hybrid soft computing methods is compared with that of corresponding single models and the results are discussed.

Chapter 6 provides the conclusions drawn from the study and highlights the important findings. A report on limitations of the study and future scope is also included in this chapter.

CHAPTER 2

LITERATURE REVIEW

2.1 General

Characterized by high complexity, dynamism and non-stationarity, hydrological and hydro climatologic forecasting has always presented a challenge to hydrologists who recognize its essential role in environmental and water resources management as well as in water-related disaster mitigation. Recent years have seen a significant rise in the number of scientific approaches applied to hydrologic modeling and forecasting, including the particularly popular ‘data-based’ or ‘data-driven’ approaches. Such modeling approaches involve mathematical equations drawn not from the physical process in the watershed but from an analysis of concurrent input and output time series (Solomatine and Ostfeld 2008). Such models can be defined on the basis of connections between the system state variables (input, internal and output variables) with only a limited number of assumptions being made regarding the physical behavior of the system (Nourani et al. 2014).

Typical examples of data-driven models are rating curves, the unit hydrograph method and various statistical models (Linear Regression; LR, multi-linear, Auto-Regressive Integrated Moving Average; ARIMA) and methods of machine learning. The conventional black box time series models such as ARIMA, ARIMA with exogenous input (ARIMAX) and Multiple Linear Regression (MLR) are linear models and assume stationarity of the dataset. Such models are unable to handle non-stationarity and non-linearity involved in hydrological processes. As a result, many researchers have focused on developing models that are able to model non-linear and non-stationary processes (Nourani et al. 2014).

The data-driven methods of Artificial Intelligence (AI) have shown promise in modeling and forecasting non-linear hydrological processes and in handling large amounts of dynamicity and noise concealed in datasets. Such properties of AI-based models are well suited to hydrological modeling problems. Numerous AI tools or techniques have been used, including versions of search optimization, mathematical optimization, as well as logic, classification, statistical learning and probability-based methods (Luger2005; Nourani et al. 2014).

Despite the flexibility and usefulness of AI-based methods in modeling hydrological processes, they have some drawbacks with highly non-stationary responses, i.e., which vary over a wide scale of frequencies, from hourly to multi-decadal. In such instances of ‘seasonality’, a lack of input/output data pre/post-processing, may not allow AI models to adequately handle non-stationary data. Here, hybrid models which combine data pre/post-processing schemes with AI techniques can play an important role.

Hybrid hydrological models may take advantage of black box (here AI-based) models and their ability to efficiently describe observed data in statistical terms, as well as other prior information, concealed in observed records (Nourani et al. 2014).

Recently, hybrid systems which performs better compared to conventional counterparts e.g., the integration of artificial neural networks with conceptual models (Chen and Adams 2006), wavelet and neuro-fuzzy conjunction model (Shiri and Kisi 2010), ANFIS (Adaptive Neuro-Fuzzy Inference System) (Tayfur and Brocca 2015) or hybrid intelligent systems (Bhadra et al. 2010) has been remarked. The wavelet-based seasonal models are more efficient than only Autoregressive models (i.e., ANN and ANFIS) for representing peak values (Nourani et al. 2014).

2.2 Hydrologic time series and their characteristics

A data series or time series is nothing but a sequential order of observed data values of a financial or physical variable taken at regular time periods Δt, symbolized as a discrete set of values X_1, X_2, X_3, ...X_n, *etc*. In the field of engineering, the series of data values is acquired by the sampling of associated continuous data from the sensors. Usually, time series data is based on observed values and associated with noise component, contains a deterministic as well as a stochastic component indicating the noise intervention that roots the statistical variations in the deterministic component values.

The investigation of a particular time series is basically intended at examining its inner structure which includes trend, autocorrelation, seasonality, *etc.*, to achieve an enhanced learning of the underlying dynamic processes by which the data values of time series are produced. The data forecasts of the time series support the decision making of the forthcoming actions of control in the process control systems.

Naturally occurring time series which includes climatic, hydrologic, and environmental data time series influence the assumptions of randomness, homogeneity, non-periodic, stationarity and non-persistence appear to be the fairly an exception (Rao et al. 2003). All the studies in water resources that make use of hydrologic time series data require preliminary step of statistical data analyses to verify whether the data series considered passes all the necessary characteristics/assumptions (Adeloye and Montaseri 2002). The detailed study of the literature on the hydrologic time series analysis so far (Machiwal and Jha 2006) exposed that there are no studies which accounted all the characteristics of time series data analysis. Majority of the study reported on considering only the linear trend analysis, and the characteristics such as stationarity, homogeneity, persistence and periodicity have been neglected in the hydrologic time series analysis. Limited number of studies can be found till date that concerns a comprehensive analysis of above-mentioned characteristics of hydrologic time series (Machiwal and Jha 2012).

2.2.1 Characteristics of hydrologic time series (Machiwal and Jha 2012)

1. Homogeneity involves all the composed data of hydrologic time series which fall under one statistical population that has mean value invariant of time. As a result, the methods to verify this property of the data series stand upon evaluation of the importance of variation in the "mean" value. Three tests for homogeneity namely, Cumulative Deviations, Bayesian test and the von Neumann test are popularly known.
2. A data time series is called as *firmly stationary*, if there are no variations in its statistical properties with changes in the time origin. For instance, if two-time intervals (non-overlapping) are chosen from a particular time series, both the sub-series will appear nearly identical. The actual subseries will be different from each another, but they will be spread around the same mean value. Consequently, a time series which is said to be stationary can never have any

kind of periodic or trend component. Due to this reason sometimes tests of periodicity and trend are employed to ensure the stationarity property of hydrologic time series. In general, there are two approaches for examining the stationarity: *parametric approach* and *nonparametric approach*. Both the approaches are used in hydrology. The two tests, viz., t-test and Mann-Whitney test are found in literature to examine the stationarity property of hydrologic time series.

3. Trend component is one general deterministic component in any time series. The tendency for consecutive values to be either increasing or decreasing over a period of time is called trend (Haan 2002). The variations in any hydrologic conditions either by natural and/or by artificial aspects can bring in the components of linear or nonlinear trends within the hydrologic time series. This component in any time series can be examined by an appropriate linear or nonlinear model. The linear model which is broadly used in hydrology is the student's t-test or a non-parametric test like Mann-Kendall's test.
4. The property like "Periodicity" in any hydrologic data time series can be identified if the data series are obtained at less than one-year intervals of time. In most of the cases, six and 12 months of periodicity are very frequent. The method of Fourier transform series has been generally employed for the recognition of components of periodicity in the hydrologic data time series.
5. The characteristic called "Persistence" can occasionally consider being periodicity. Many of the hydrologic data time series researches show that there is no difference between randomness and persistence. Hence, the tests to study the property of randomness in a hydrologic data time series are usually employed for checking both persistence and trend (Machiwal and Jha 2012).

2.3 Applications of soft computing techniques in hydrologic time series forecasting

Accurate predictions of streamflow values are crucial for increasing the efficiency of reservoir operations, networks in water supply, river flood mitigation and management of water resource systems. In recent years, data-driven or AI models like artificial neural networks (ANN) and adaptive neuro-fuzzy inference system (ANFIS), have been used as efficient methods for modelling of complicated and nonlinear systems in hydrology (Figueiredo 2007; Seo and Kim 2016). Although ANFIS and

ANNs are extensively useful for forecasting of hydrological parameters, they too have some issues while handling the non-stationary series of data (Seo et al. 2015; Seo and Kim 2016). Generally, hydrological time series data contains a number of frequency components and are associated with nonlinearity, a variety of hybrid combinations of modelling techniques have been introduced and employed for the advancement and improvement of model performance and the forecasting ability (Okkan 2012; Seo and Kim 2016). Particularly, the combination of wavelet decomposition method and data-driven or AI models has proved to be successful and efficient in modelling and forecasting hydrological variables (Seo and Kim 2016).

Budu et al. (2012) proposed a novel hybrid modeling approach where they combined wavelet transform and the Artificial Neural Networks (ANN) and developed wavelet neural network (WNN) model. The novel model was applied for modeling of river flow time series. The hybrid WNN model was capable of performing well, particularly the extreme values during the model's testing period. The performance of WNN model in applications implied that the hybrid modeling approach yielded enhanced results in extracting the properties of hydrograph compared to the single models (ANN and AR).

Santos and Silva (2014) introduced a hybrid wavelet and artificial neural network (WA)model for forecasting daily streamflow data and found that the hybrid models prove improved performance when compared to classical ANN models. Mehr et al. (2013) compared Feed-forward-neural-network (FFNN) model and the Neuro-Wavelet (NW) model to forecast monthly streamflow data where they used the wavelet function to decompose the time series in to sub series.

Keskin et al. (2006) used fuzzy logic approach to forecast monthly flow values and found that it gave suitable results with historic flow values. They set up a fuzzy flow estimation model between two observation stations in a stream to predict the downstream flow values from the upstream flow values. Fuzzy model gave better results compared to the rational method.

Recently, the wavelet transforms theory and applications have been initiated in the hydrological domain. The theory of wavelet analysis has been recognized in recent times as one of the useful techniques for unfolding both runoff and rainfall data time-series. Wavelet is used as an efficient tool for reduce noise in the hydrologic time series. A detailed study on “noise” and its effects on forecasting can be found in Elshorbagy et al. 2002. The Wavelet-Autoregressive model (WARM) is developed for annual rainfall prediction by combining the wavelet transform method with the conventional AR model. Coulibaly and Burn (2004) applied the wavelet method of analysis to recognize and illustrate unpredictability in Canadian annual stream flows and to study the dynamic linkage between the stream flows and the predominant variations of climate parameters in the Northern Hemisphere. Owing to the resemblance between wavelet decomposition method and one hidden neural net-work layer, the concept of merging both wavelet and neural network has made the combination of wavelet neural network possible, which has been applied in several fields. Performance of such hybrid model showed that, the adaptation and training competence of the wavelet neural network is superior compared to other networks. Dongjieet al. (2004) applied a conjunction of neural networks and wavelet transforms techniques to forecast levels of ground water. Aussem and Murtagh (1997) employed a Dynamical Recurrent Neural Network (DRNN) for every resolution level of the sunspot time series which result from the decomposed wavelet series with the Temporal Recurrent Back propagation (TRBP) learning algorithm. Partal (2007) developed a model in conjunction of wavelet-neuro-fuzzy to predict the daily precipitation of Turkey region. The observed daily precipitation values are decomposed in to some sub series by Discrete Wavelet Transform (DWT) and then the suitable sub series are considered as inputs to neuro-fuzzy models for forecasting the daily precipitation series.

2.4 Various hybrid fuzzy models in hydrologic time series forecasting

Table 2.1 Applications of hybrid fuzzy model in hydrologic time series forecasting

Hybrid models	**Sl. No.**	**Author/Authors**	**Applications**	**Data used/models developed/results obtained**
Neuro-fuzzy/ AdaptiveNeuro-Fuzzy Inference System (ANFIS)/Fuzzy neural networks	1	Deka and Chandramouli (2003)	River stage-discharge relationship	Compared four methods like neural network (NN) model, modularized NN model, conventional curve fitting method and a fuzzy NN model. Fuzzy NN model produced best results in the study.
	2	Bae et al. (2007)	Forecasting dam inflow	They used ANFIS model to forecast the optimal dam inflow. Past observed data and

				weather forecasting information were used for development of the model.
	3	Deka and Chandramouli (2009)	Reservoir operation	Developed Fuzzy Neural Network (FNN) model to study the optimal operating of a reservoir. They studied the advantages of FNN model over Dynamic programming.
	4	Pramanik and Panda (2009)	River flow prediction	Artificial Neural Networks (ANN) and ANFIS models were developed to estimate the discharge at the downstream of a river. Comparison of the models was done by estimating the discharge from a barrage at downstream. Results of ANFIS were closer to the observed discharge and hence it functioned better than ANN model.
	5	Talei et al. (2010)	R-R modelling	Applied ANFIS model in event-based R-R modeling. ANFIS model results were compared with conventionalstormwater management model (SWMM). ANFIS was found to be better at estimating peak flow compared to SWMM.
	6	Jeong et al. (2012)	Forecasting of monthly precipitation	Applied ANFIS model to forecast qualitative and quantitative monthly precipitation. Results showed that ANFIS can be a promising approach for forecasting qualitative monthly precipitation.
	7	Talei et al. (2013)	Runoff forecasting	Applied Neuro-Fuzzy system (NFS) for R-R modeling. NFS was compared with three other hydrologic models in order to prove its efficiency.
	8	Chang et al. (2014)	Forecasting of watershed rainfall	Used ANFIS model for predicting watershed rainfall, which served as a valuable data for flood warning system during periods of the typhoon.
Fuzzy logic with Support Vector Machine (SVM)	1	He et al. (2014)	River flow prediction	Three potential methods ANN, ANFIS, and SVM were used for forecasting river flow. SVM model performed better than other two models.

Fuzzy logic with Wavelet model	1	Partal and Kisi (2007)	Precipitation forecasting	Combined wavelet and neuro-fuzzy (NF) models to develop Wavelet-Neuro-Fuzzy (WNF) model to predict precipitation. WNF model produced significantly better outcomes compared to classical neuro-fuzzy models.
	2	Ozger et al. (2012)	Drought forecasting	This study combined wavelet and fuzzy logic to produce wavelet fuzzy logic (WFL) model to forecast long- lead time droughts. WFL model results were more accurate for drought forecasting compared to ANN and coupled wavelet and ANN (WANN) models.
	3	Sahay and Sehgal (2014)	Forecasting monsoon flows	Wavelet-ANFIS (WANFIS) model was developed to forecast current-day flow in a river when provided with only historical flow data. WANFIS showed high accuracy compared to ANFIS and auto-regression (AR) models.
Fuzzy logic with Genetic programming (GP)	1	Han et al. (2012)	Reservoir operation	Fuzzy programming and a self-adaptive GA were used for eco-friendly reservoir operation. The presented methodology showed potential applications in reservoir operation.
	2	Young et al. (2015)	Forecasting of watershed runoff	This study utilized three model approaches for predicting runoff. The hydrological engineering center hydrological modeling system (HEC-HMS) was combined with two hybrid models: Genetic Algorithm Neural Network (GANN) and ANFIS. Both models performed significantly well in improving the prediction accuracy.

2.5 Stream water quality-quantity modeling

Fuzzy theory is one of the powerful techniques with the potentiality of resolving several complicated problems in hydrology which also involves the assessment of stream water quality along with handling the uncertainties associated with the data and ambiguity that arise in a river or a stream system. Fuzzy techniques can be very helpful in handling the random nature of hydrologic parameters and uncertainties that arise from missing data problems. The Fuzzy interference systems (FIS) play a significant character in fuzzy implementations. The performance of FIS can essentially be enhanced by using other techniques like “grey clustering” methods and “similarity measures” in order to achieve superior and precise results for water quality assessment. The FIS systems can also be incorporated with expert systems and efficient decision support systems to support the decision-makers in enhancing stream water quality by employing efficient approaches and measures. Therefore, the ability of fuzzy systems has to be integrated with other efficient methods such as artificial neural network, expert system and grey clustering techniques to offer a wide-range of solutions to prevent and manage stream water pollution problems was studied by Oladipo et al. (2021).

A novel fuzzy system optimization method was introduced for the purpose of managing the seasonal water quality of river systems. This model addressed the uncertainty involved in a water quality system by using a fuzzy probabilistic framework. The probability of occurrence of an event of poor water quality was considered as a “fuzzy event”. The associated randomness with the index for water quality was connected to this “fuzzy event” by the idea of probability of occurrence of a fuzzy event (Mujumdar and Sasikumar 2002).

Leelavathy et al. (2016) utilised a Fuzzy Inference System (FIS) in an interesting way to evaluate the Water Quality Index (WQI). They examined the effects of the several variables such as biological oxygen demand (BOD), Dissolved oxygen (DO), Total Coliforms (Tco), Temperature (Temp) and pH on the desire driver water quality. They developed two fuzzy inference systems, with parameters Tco, DO and BOD as inputs for one system and the other system with the results of first FIS, Temp and pH.

Management of stream water quality issues have been studied as multi-objective optimization problems in the literature so far. Also, management of water quality issues are considered by vagueness in their water quality standards and objectives. Fuzzy optimization and fuzzy sets present a constructive method in dealing such ambiguity. The imprecision in model parameters and objective functions cause uncertainties in water resources problems which can be modelled and handled with fuzzy sets.

LI Ru-zhong (2007) developed a two-dimensional fuzzy model for water quality where they defined the quality parameters as symmetrical triangular fuzzy numbers and for the application in unexpected pollutant discharge. The pollutant concentrations corresponding to a particular confidence level were found with the help of the “α-cut” technique and some of the arithmetic operations performed on the triangular fuzzy numbers. This study demonstrated the feasibility of triangular fuzzy numbers both in theory and calculations to the simulation of stream water quality parameters.

A few researches can be seen in literature that studied the connection between the flux rates, discharge and nutrient concentration (Bhat et al. 2015). Bodo and Unny (1983) verified with case studies and showed the importance of variations of flow which is a basic transport system for sediment transport and that there is a necessity to model such flow events exclusively so as to accomplish consistent sediment load estimations. The study concluded that minute flashy river streams need to be sampled during flow events for a greater intensity than the bigger river streams with comparatively attenuated response of flows so as to generate suspended sediment load estimations of reasonable accuracy. Robertson and Roerisch (1999) demonstrated that the samples collected during high-flows, typically to assist in defining the connection between elevated stream flows and the corresponding heavy loads, consequences in inaccuracy, overvalued annual sediment loads if such samples are constantly obtained during near the beginning of high-flow rate events.

Vanni et al. (2001) showed that most of the sediment exports occur during high storm flows for all fractions of the nutrients and therefore, the nutrient transport highly relies on the hydrological regimes. In 2001, Clement (2001) anticipated yearly nutrient loads transport to Lake Balaton for over three decades where he demonstrated that about 50% of the loads are transported in to the lake by its tributaries. The study also assessed the uncertainty degree with the help of sampling theory and obtained load-flow relationships.

The study by Vieux and Moreda (2003) studied the relationship between discharge rates and nutrient concentrations and demonstrated that greater part of annual phosphorus (P) loads is carried by the direct surface runoff where heavy concentrations of loads carried by high flow rates and lighter loads by low flow rates.

In addition, for rivers and lakes of short to medium time scales of hydraulic residence, the major share of nutrient input is carried and added by their watersheds. Therefore, management at watershed-scale has demonstrated to be proficient for the reduction of water pollution which essentially involves an assessment of the pollutants loading at the outlet of the watershed (Schindler 1978; Guo et al. 2002).

2.6 Summary literature review

The review provided a detailed description of various soft computing methods used for hydrologic forecasting. The literature review stressed on the importance of data pre-processing in hydrologic forecasting especially in streamflow forecasting. The idea of hybrid modeling was derived when single models showed their limitations for a particular application. Two or three soft computing models were combined in order to eliminate the limitations of each other and hence improved the overall performance.

The concept of linking the stream water quality and quantity is not available in literature so far (especially linking the quality parameters with the quantity of flow at that particular period of time). Although there are papers related to water quality index and other water pollution measuring methods, literature related to soft computing to determine water quality index are very few. The application of fuzzy logic model to develop water quality index is relatively new in the field of hydrology and water resources.

Most of the past works related to hybrid wavelet-AI models used to forecast time series have focused on conventional AI models such as AR models and ANNs (Peng et al. 2017; Santos and Silva 2014; Honorato et al. 2018; Chong et al. 2019). In the present study, we have made an attempt to compare the performance of various hybrid wavelet-AI methods such as SVM, ANFIS and Fuzzy logic, which are comparatively new in the field of stremflow forecasting.

2.7 Motivation for this research

- From literature survey, it is evident that most of the research on wavelets is limited to denoising using a few mother wavelet functions. Many studies lack in determining the best suitable mother wavelet and the variations in the results with change in mother wavelets. In our study, we have performed the sensitivity analysis of various mother wavelet functions and their suitability for forecasting application.

- Hydrologic time series forecasting using soft computing tools has been experimented by many researchers so far. But combinations of hybrid fuzzy modeling are not exhaustively used in literature.
- Various combinations of hybrid soft computing techniques like wavelet-fuzzy, fuzzy genetic algorithms and fuzzy-SVM are comparatively new and innovative in case of hydrologic modeling time series.
- There is a need to improve the accuracy of hydrologic time series forecasting models by considering the various hybrid combinations of soft computing techniques.
- Wavelet-Fuzzy based combination is a strong hybrid technique which uses the decomposing nature of wavelets and the linguistic rules of fuzzy logic to deal with uncertainty in hydrologic time series modeling.
- The research that relates stream water quality and quantity is comparatively new in the field of hydrology and water resources.

2.8 Problem formulation

It is necessary to provide a powerful tool to reduce the noise in the data so that accuracy of the model is increased. The noise is that part of the series which includes the odd values in the series such as missing data, mis-interrupted values and unexpected high peaks (Badrzadeh et al. 2018). Wavelet transformer is one such powerful tool used to decompose the data set into different scales (Surendra and Deka 2015). In this research work, discrete wavelet transform is coupled with the fuzzy logic method to improve the accuracy of the forecast.

The second phase of this research is stream water quality-quantity modeling. Water quality and quantity are the two aspects that are interrelated and hence should be studied together within an integrated framework. In today's world, demand for water essentially takes into account both quality and quantity aspects for various uses of water (Ganoulis 2009). Having a sufficient accessible quantity of water becomes meaningful only if this quantity of water is acceptable in terms of its quality (Ahmed et al. 2015). This study aims at studying the role of the quantity of water in determining its quality along with the other quality parameters.

2.9 Objectives of the study

The primary objectives of the study are as follows:

1. To understand the importance of data pre-processing in hydrologic time series forecasting.
2. To investigate the strength of various mother wavelet functions for de-noising the streamflow time series.
3. To evaluate and compare the performance of hybrid wavelet-soft computing techniques like Fuzzy logic (WT-Fuzzy), ANFIS (WT-ANFIS), and SVM (WT-SVM) for daily streamflow forecasting.
4. To examine the relationship between quantity and quality of stream water.

CHAPTER 3

METHODOLOGY

3.1 General

The major steps implemented in the study are as under:

Part 1: Forecasting of streamflow using wavelet coupled model

Step 1: Forecasting of streamflow series by using wavelet coupled soft computing approach.

Step 2: Comparing the results of single Fuzzy, ANFIS and SVM models with the respective wavelet-hybrid models.

Step 3: To examine the improvement in the results when coupled with wavelet transform.

Part 2: Stream quality modeling

Step 4: Examining the relationship between stream water quality-quantity.

3.2 Part 1: Forecasting of streamflow using wavelet coupled model

Streamflow forecasting is essential in many activities involving the operation and optimization of water resources. For this reason, the development of mathematical models able to provide more reliable long-term forecasting has attracted the attention of hydrologists through time.

3.2.1 Modeling strategy

Workflow of the study and flow chart of methodology adopted in the present study for streamflow forecasting is as shown in Figure 3.1. In the present study, the number of lags is considered based on ACF and PACF functions. The present day and the future forecasts of streamflow essentially depend on its historical or past records. Hence it is important to consider the streamflow data with respective number of lags determined by ACF and PACF functions. Number of lags represent the period of past data which influence the forecasts. The raw streamflow data is processed using DWT with various mother wavelets decomposed till level 5. For level 5 decomposition, one approximation and 5 detail coefficients are obtained. The approximations and details at each level are fed to the SVM, ANFIS and Fuzzy models as inputs. The mother wavelet which shows good performance is selected as the best model for each station to forecast the 1-day, 3-days and 5-days ahead streamflow for that particular gauging station.

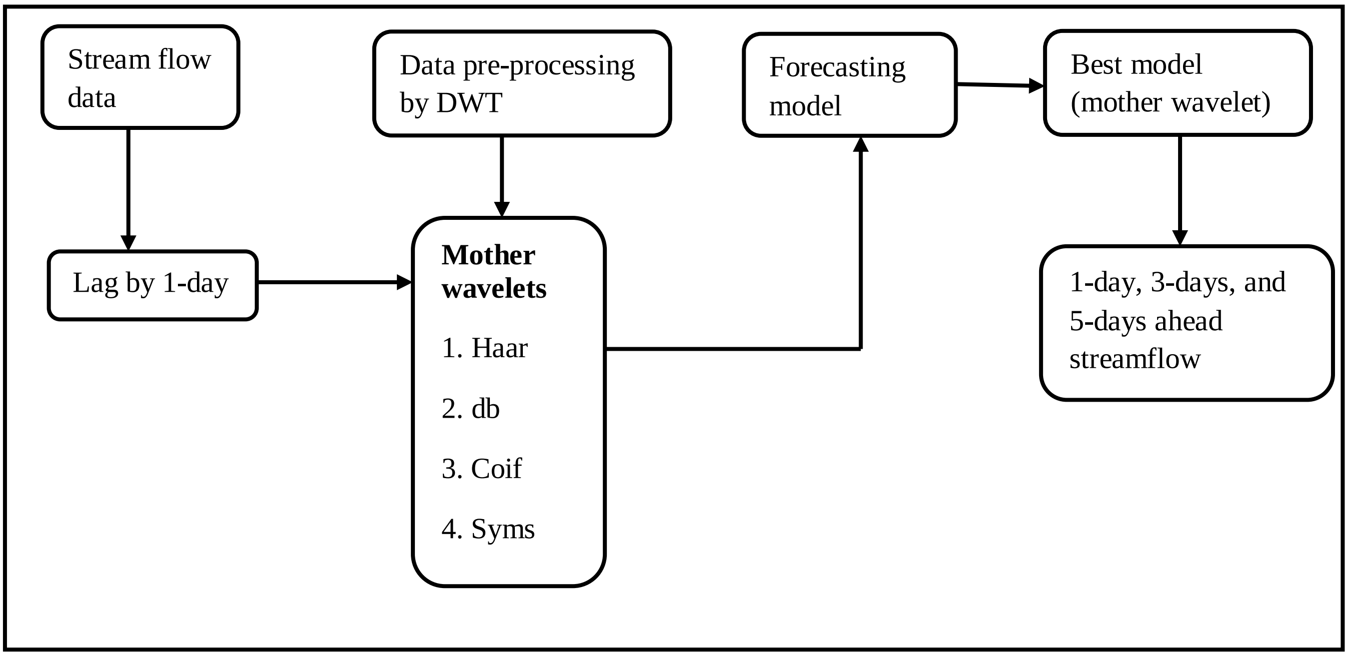


Figure 3.1 Methodology illustrated through a series of steps for streamflow forecasting

Choosing a proper lead time is crucial, especially for daily time series forecasting as the accuracy of models gets affected for longer lead times (Nourani et al. 2014b). Adamowski (2008) stated that wavelet conjunction models are accurate for short term forecasting (1 and 3 days) and less accurate for long term forecasting (7 days and more) for stream flow as well as flood forecasting (Hadi and Tombul 2018). In the present study, we have considered 1 day, 3 days and 5 days lead time forecasts produced by single forecasting model and examined the improvement in the results of those forecasts when wavelet pre-processing is applied.

3.2.2 Data pre-processing using Wavelet Analysis

A wavelet is a function used to localize a series in both space and scaling. Using the mother wavelet $\psi(t)$, a family of wavelets can be synthesized (Torrence and Compo 1998). There are two types of wavelet analysis: continuous wavelets transform (CWT) and discrete wavelets transform (DWT). Sang et al. (2016) reported in their work that

DWT method has got more applications compared to CWT in the field of hydrologic time series forecasting for the reason that the hydrological data is usually measured and noted in discrete time steps (Adamowski and Sun 2010; Tiwari and Chatterjee 2010).

The CWT method suffers from data redundancy problem when large amount of data is fed to the system, whereas, DWT method overcomes this problem by using orthogonal wavelets. Hence in this study, we have used orthogonal mother wavelets like Haar, Daubechies (db), Coiflets (coif) and Symlets (sym). Another advantage of DWT method is that it avoids overlapping of multi-components and the influence of noise in wavelet coupled artificial intelligence hybrid modeling (Sang et al. 2016). Hence, DWT is best suitable and commonly used method for hydrological time series forecasting.

DWT uses wavelet function and scaling function known as high pass filter and low pass filter respectively. The DWT in the mathematical form is defined by (Hadi and Tombul 2018):

$$Wf(j,k) = \int_{-\infty}^{+\infty} f(t)\psi_{j,k}^{*}(t)dt \tag{3.1}$$

$$\psi_{j,k}(t) = a_0^{-j/2}\psi(a_0^{-j}t - b_0 k) \tag{3.2}$$

where a_0 and b_0 are constants, j is the decomposition level, and k is the time translation factor. The DWT frequently used is the dyadic DWT obtained by tuning the values $a_0 = 2$ and $b_0 = 1$ and that directs to the function of Equation 3.3 (Daubechies 1992):

$$\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k) \tag{3.3}$$

At first, on using a dyadic DWT, the signal is passed through the filters resulting in two types of coefficients sets: approximation produced by low pass filter, and detail produced by high pass filter. At each subsequent level, say i^{th} level, one approximation and i number of detail coefficients are produced. Here in the present study, we have considered the 5^{th} level of decomposition as the optimum level by trial-and-error method.

3.2.3 Streamflow forecasting Fuzzy logic, ANFIS and SVM models

3.2.3.1 Fuzzy Logic (FL)

In order to implement the technique of fuzzy logic in to a real-life application, one need to follow three steps (Figure3.2):

1. Fuzzification – conversion of classical/crisp data into fuzzy data (Linguistic) or Membership Functions (MFs)

2. Fuzzy Inference Process – combination of membership functions with the if-then or control rules to derive the fuzzy output

3. Defuzzification – Usages of different methods to evaluate each associated output and organize them into a table called the lookup table. Then pick up the output from the lookup table based on the current input based on the application

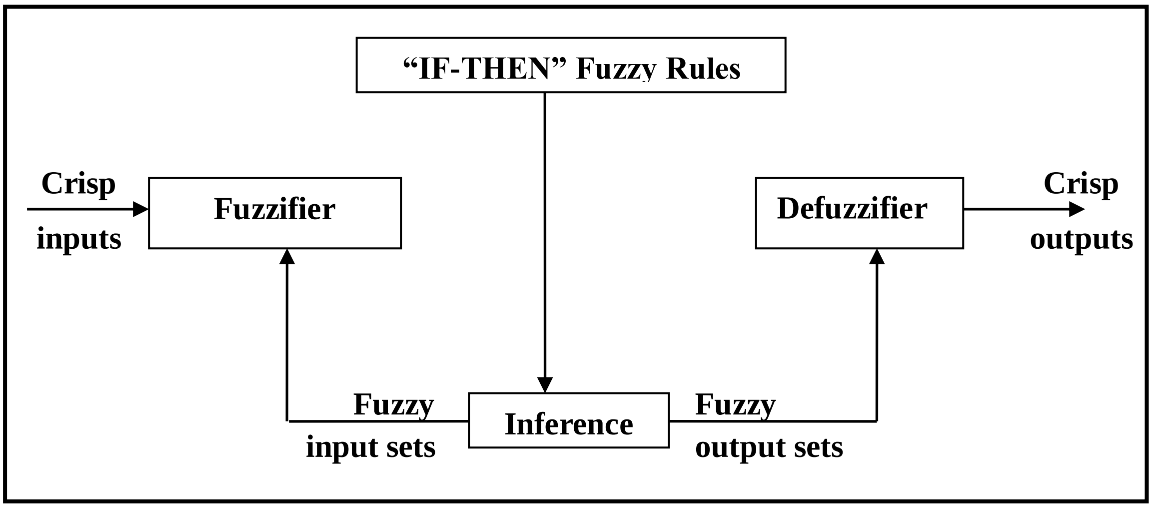


Figure 3.2 The technique of Fuzzy logic

3.2.3.2 ANFIS

Jang (1993) was the first to propose Adaptive Neuro-Fuzzy Inference System (ANFIS) which was later elaborated by Jang et al. (1997) as a universal function estimator to evaluate any factual continuous functions. Basically, the FIS is categorized into three types, namely, the Mamdani inference system defined by Mamdani and Assilian (1975), the Tsukamoto inference system by Tsukamoto (1979), and the Sugeno inference system by Takagi and Sugeno (1985) which are based on the operation of inference used in the “if-then rule”. In ANFIS, the Sugeno type of inference system is used.

The basic structure of the learning algorithm adopted in ANFIS consists of a hybrid combination of back-propagation, gradient descent and least-squares methods. The ANFIS system can be clarified by using two inputs (say, ×1 and ×2) and one output (say, y). The former may symbolize the flow of two predecessor days say Q_{t-1} and Q_{t-2}, while the latter symbolizes the present-day flow Q_t. Using the above denotations, two typically formed “if-then rules” of the first order Sugeno fuzzy model can begiven as follows:

Rule1: “If x_1 is A_1 and x_2 is B_1, then $f_1 = p_1x_1 + q_1x_2 + r_1$”,
Rule2: “If x_2 is A_2 and x_2 is B_2, then $f_2 = p_2x_1 + q_2x_2 + r_2$”,

here p_1, q_1, r_1 and p_2, q_2, r_2, are the variables in the sub-sections of the ANFIS model, and A_i and B_i are the corresponding linguistic labels provided by the membership function. The fuzzy rules related two or more input variables to the single or multiple outputs using “if-then” rules. For example, “if” temp is X^0C “and” humidity is Y%, “then” rainfall is Z mm. A detail elaboration of ANFIS structure can be referred in Kişi et al. (2012).

3.2.3.3 SVM

This study utilizes SVM as a forecasting model for daily stream flow forecasting. The main advantage of support vector machine is that it uses the kernel trick in order to build expert knowledge so as to reduce model complexity.

The regression function of SVM in constrained form (Raghavendra & Deka, 2014) is given as follows:

$$Minimize\colon \frac{1}{2}\|\omega\|^2 + C\left(\sum_{i}^{N}\left(\xi_i + \xi_i^*\right)\right) \tag{3.4}$$

$$Subject\ to \begin{cases} \omega_i\phi(x_i) + b_i - d_i \le \varepsilon + \xi_i^*, i = 1,2,\dots,N & (3.5) \\ d_i - \omega_i\phi(x_i) - b_i \le \varepsilon + \xi_i, i = 1,2,\dots,N & (3.6) \\ \xi_i, \xi_i^* \ge 0, i = 1,2,\dots,N & (3.7) \end{cases}$$

where$\phi(x_i)$ is the aspect of inputs, ω_i and b are the coefficients, C is the regularized constant, ε is the penalty or error and ξ andξ* are the slack variables.

By introducing Lagrange multipliers and applying Karush-Kuhn-Tucker (Fletcher 1987) conditions to the regression Equation (3.4), the dual Lagrangian equation takes the following form:

$$v(\alpha_i, \alpha_i^*) = \sum_{i=1}^{N} d_i\,(\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^{N} (\alpha_i + \alpha_i^*) - \frac{1}{2} \sum_{i=1}^{N} \sum_{i=1}^{N} (\alpha_i - \alpha_i^*)(\alpha_i - \alpha_i^*) K(x, x_i) \qquad (3.8)$$

along with the constraints as,

$$\sum_{i=1}^{N} (\alpha_i - \alpha_i^*) = 0 \; and \alpha_i, \alpha_i^* \in [0, C], i = 1,2, \dots, N) \qquad (3.6)$$

where $\alpha_i$and $\alpha_i^*$are the Lagrangian multipliers and K (x, x_i) is known as kernel function. In this study, we have used Radial basis kernel function (RBF) (Kisi and Cimen 2011):

$$K(x, x_j) = \exp(-\|x - x_j\|^2 / 2\sigma^2) \qquad (3.9)$$

where σ is the Gaussian noise level of standard deviation.

SVM model for regression can be most efficient if kernel parameters, regularization parameter C and ε-insensitive tube are tuned properly. In this study, the quadratic optimization is done by using sequential minimal optimization (SMO) technique in MATLAB interface. Platt 1998 proposed a new algorithm for training support vector machines: Sequential Minimal Optimization, or SMO. Training a support vector machine require the solution of a very large quadratic programming (QP) optimization problem. SMO breaks this large QP problem into a series of smallest possible QP problems. These small QP problems are solved analytically, which avoids using a time-consuming numerical QP optimization as an inner loop. The amount of memory required for SMO is linear in the training set size, which allows SMO to handle very large training sets.

3.2.4 Performance Evaluation Methods

In this study, three performance indices have been utilized to evaluate the performance of the models. They are coefficient of determination (R^2), root mean square error (RMSE) and Nash-Sutcliffe Efficiency (NSE) (Nash and Sutcliffe 1970). Equations used to compute these measures are listed in Table 3.1.

Table 3.1 Performance indices used in the study

Performance index (PI)	Formula	Definitions
Coefficient of determination (R^2)	$\frac{[\sum_{i=1}^{N}\left(Q_0(i)-Q_0m(i)\right)*(Q_f(i)-Q_fm(i))]^2}{\sum_{i=1}^{N}\left(Q_0(i)-Q_0m(i)\right)^2*\sum_{i=1}^{N}(Q_f(i)-Q_fm(i))^2}$	$Q_0(i)$= Observed streamflow $Q_f(i)$ = Forecasted streamflow $Q_0m(i)$ = mean of observed $Q_fm(i)$ = mean of forecasted
Root means square error (RMSE)	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(Q_f(i)-Q_0(i))^2}$	
Nash-Sutcliffe Efficiency (NSE)	$1-\frac{\sum_{i=1}^{N}(Q_f(i)-Q_0(i))^2}{\sum_{i=1}^{N}(Q_0(i)-Q_0m(i))^2}$	

N: number of data; PI range: $0 \le R^2 \le 1$; $0 \le$ RMSE $\le \infty$; $-\infty \le$ NSE ≤ 1.

3.3 Part 2: Stream quality modeling

The water quality parameters like Dissolved Oxygen (DO), pH, Electric Conductivity, Biochemical Oxygen Demand (BOD), Nitrogen as nitrate (Nitrate-N), Fecal Coliform (FC), Total Coliform (TC), Nitrogen as ammonia (NH4-N), Chemical Oxygen Demand (COD), Total Hardness as CaCO3 (TH as CaCO3) and Total Alkali as CaCO3 (TA as CaCO3) are briefly explained below.

The basic quality parameter, pH is used to quantify the alkalinity or acidity of water. Also, pH is a vital factor for limiting and for sustainability of domestic as well as aquatic life. It is usually expressed as a number in a scale ranging from 1 to 14. The pH of pure distilled water is 7. Under base flow conditions, pH of the stream water largely depends on the underlying soil properties and geology of the watershed. The natural process of photosynthesis by plants and algae in the aquatic ecosystems can significantly cause variations in its pH. Another important cause for pH variations is the contaminations from industrial wastewaters and storm waters. The changes can create either alkaline or acidic natures in the river ecosystems. An acceptable range of pH for freshwater is 6.5 – 8 and 8 – 9 is the optimal range for sea waters and estuaries (Subramani and Ismail 2015).

The overall health of water is represented by the amount of oxygen present in it. Dissolved Oxygen (DO) is expressed in terms milligrams per litre and is a measure of total quantity of oxygen present in the water. The saturation of DO is very crucial for the organisms in aquatic ecosystems. The decomposition of decayed aquatic plants, effluents from sewage disposal, storm water discharge contaminations and wastewater discharges from man-made activities, by the micro-organisms, all eventually lessen the DO levels in the river water. River waters that have sufficient levels of DO present, can generally support diverse aquatic ecosystems. The DO level of 3 mg/L is generally considered as suffocating to most of the aquatic lives. Any water with as low DO as 0.5-2 mg/L is said to be hypoxic and those with DO less than 0.5 mg/L are called as anoxic (Sahoo 2014).

The quantification of total quantity of oxygen utilized by the biological and chemical activities in the river over a period of 5 days is measured as its BOD. It is determined by quantifying the present oxygen level in the water sample collected and then measuring the DO after storing it for 5 days in the dark at a steady temperature condition of 200°C. The basic difference between DO and BOD is that the demand or consumption of oxygen by chemical and biological activities in the river. Usually, BOD is measured in terms of milligram per litre of water. In general, unpolluted and natural stream waters must possess a BOD of less than or equal to 5 mg/L. The BOD of raw sewage may vary from 150-300 mg/L (Subramani and Ismail 2015).

The capability of water to pass electric current, especially when the water is influenced by the existence of certain dissolved solids like nitrate, chloride, sulphide, sodium, phosphate, magnesium, iron, aluminium and calcium. Temperature can also affect EC with higher EC for warmer waters and vice-versa. Further, higher EC values can be seen in river waters that flow across the regions with soils of clayey nature due to the presence of certain minerals that ionize when dissolved in water. EC is measured in micro-Siemens per centimetre or micro mho per centimetres. The EC for distilled water ranges from 0.5 to 3µmho/cm. The river water has EC ranging from

100 to 1000 µmho/cm. The water in streams contaminated with pollutions from industrial wastes has EC in the range as high as 10,000µmho/cm.

River water contains nitrate in minute quantities. Basically, the nitrate when dissolved in water becomes an essential nutrient for growth of plants. Industrial and manure-based fertilizers, faulty septic systems and waste water treatment plants are the most common nitrate sources. The levels nitrate concerns the health of plants, fish and other forms of life in rivers. The nitrate level for drinking water should be 45-100 mg/L (Sahoo 2014).

Total Coliforms (TC) are the pathogens which do not cause illness directly, but presence of such pathogens in the supplied water may turn to be vulnerable to harmful contaminations from microorganisms. The contamination sources of such pathogens and microorganisms are mainly due to improper treatment and discharge of septic and sewages, animal manure leaching, surface runoff/storm drainage and/or animal wastes. The measure of total coliform count is expressed by a number called Most Probable Number (MPN) usually measured per 100 ml of water. The drinking water standard for TC ranges from 50-500 MPN/100 ml.

The Fecal Coliforms (FC) are like total coliforms which are basically the basis of pathogenic contaminations or illness causing viruses and bacteria. Certain other types of non-pathogenic bacteria commonly found in intestines of animals accompany such illness causing organisms and cause diseases. The FC count or drinking purposes should range from 0-700 MPN/100ml.

The measure of total quantity of organic matter in water is called as Chemical Oxygen Demand (COD). COD can be empirically associated to organic carbon, BOD and organic matter. COD is said to be an indicator of content of organic matter present in water due to the reason that main frequent substance that is oxidized by dissolved oxygen that is there in water is the organic matter which has a biological origin such as dead plants and animal’s waste. The COD concentration is higher in the bottom of the river because of presence of higher organic matter in the bottom of the river than the surface (Subramani and Ismail 2015).

One of the highly toxic substances and most important pollutants of water in the aquatic ecosystems is the “ammonia”. It results from microbiological activities which reduce nitrogen contents in water. The major sources of ammonia are pollutions from industrial and sewage disposals and consequently, possibility of existence of pathogens and microorganisms in the water. Ammonia can be of two chemical forms in aqueous solution; one is NH4+, which is ionized form and hence less toxic and another is NH3, which is unionized and hence more toxic. The nitrogen present as ammonia in drinking water has to be in a range of 0-1.2 mg/L.

All bases in the stream water are measured in terms of total alkali and may be referred as the safeguarding ability of river water, or its capacity to resist any change in the levels of its pH. The most frequent and significant base ion is the carbonate. The total alkalinity in water is measured in terms of mg/L (milligrams per litre) of $CaCO_3$ (calcium carbonate). Neutral to moderately basic pH can be observed in waters which have high levels of total alkali. Due to the presence of bicarbonates and carbonates which neutralize or safeguard $CO_2$and acids in river waters, the pH does not change and is more stable throughout the day. The total alkali must be ranging from 0-200 mg/L as $CaCO_3$ in drinking water.

The measure of divalent cations (+2 ions) in the river water is the “total hardness” and, just like total alkalinity, it is expressed in terms of mg/L of $CaCO_3$. When dolomite and limestone and dissolve in water, calcium is one half of the molecule (the "hardness") and carbonate is the other half (the "alkalinity") of the molecule. Hence, most of the times both are equal. The concentration of calcium ions (Ca_2+) in freshwater ranges from 4 to 100 mg/L (10–250 mg/L of calcium hardness as $CaCO_3$) and that of seawater ranges from 400 mg/L (1000 mg/L of calcium hardness as $CaCO_3$) (Sahoo 2014).

3.3.1 Modeling strategy for water quality assessment

Workflow of the study and flow chart of methodology adopted in the present study for water quality-quantity assessment is as shown in Figure 3.3.

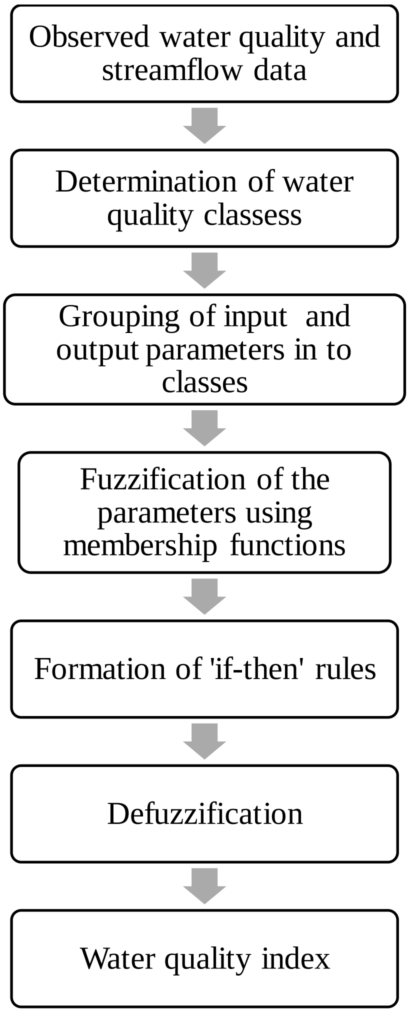


Figure 3.3 Flowchart of methodology followed in the study for water quality-quantity assessment

3.3.2 Calculation of water quality index (WQI)

The Water Quality Index (WQI) is defined as a single value determined by consideration of various parameters that are important for assessing the water quality. It is basically a combined effect of the quality parameters when taken in correct fractions for determining the water quality. WQI is usually calculated in three steps (Water programme 2007; Ramkrishnaiah et al. 2009). The procedure to assign weightages to the quality parameters were considered in order to recognize and emphasize the region-specific causes for water contamination. In the beginning, each of the quality parameter was assigned with a weight factor (wi) based on its significance relative to the overall water quality suitable for drinking, as well as relative to the percentage of collected samples falling within the permissible limits of water quality standards. The weightages such as 5, 4, 3, 2, and 1 are allotted to the water quality variables when 0-20, 21-40, 41-60, 61-80 and 81-100 % of samples respectively fall within the permissible limits (Raychaudhuri et al. 2011). Next, the relative weight (Wi) is calculated from the Equation 3.10:

$$Wi = \frac{wi}{\sum_{i=1}^{n} wi} \tag{3.10}$$

where, Wi is the relative weight, wi is the weight of each quality parameter and n is the total number of parameters.

Third step engages the assignment of a scale called "quality rating scale" (qi) for each of the parameter by dividing its concentration in each water sample by its particular standard as per the guidelines given by BIS, followed by multiplication with 100:

$$qi = (Ci/Si) \times 100 \tag{3.11}$$

where qi indicates the quality rating scale, Ci represents the concentration of each chemical parameter in each water sample in mg/L, and Si is the Indian drinking water or irrigation water standard for each chemical parameter in mg/L as per the guidelines of the BIS 10500, 2002 and 2004 or FAO respectively. For calculating the WQI, the SI is first calculated for each chemical parameter, which is then used to calculate the WQI as per the following Equation3.13:

$$SIi = Wi \times qi \tag{3.12}$$

$$WQI = \sum_{i=1}^{n} SIi \tag{3.13}$$

SIi denotes the sub-index of ith parameter; qi denotes the rating scale based on concentration of ith parameter and n is the total number of parameters. The computed WQI values are then classified into five classes, "excellent" "good", "poor", "very poor" and "unsuitable" as shown in Table 3.2.

Table3.2 WQI classes for drinking water (Rayachaudhari et al. 2014)

WQI Value	Water quality classification
<50	Excellent
50-100	Good
101-200	Poor
201-300	Very Poor
>300	Unsuitable

3.3.3 Fuzzy water quality model

In this study, we have used Mamdani Fuzzy inference system, with 9 water quality parameters and streamflow as inputs and fuzzy water quality index (FWQI) as output, which is shown in the following Figure 4.4.

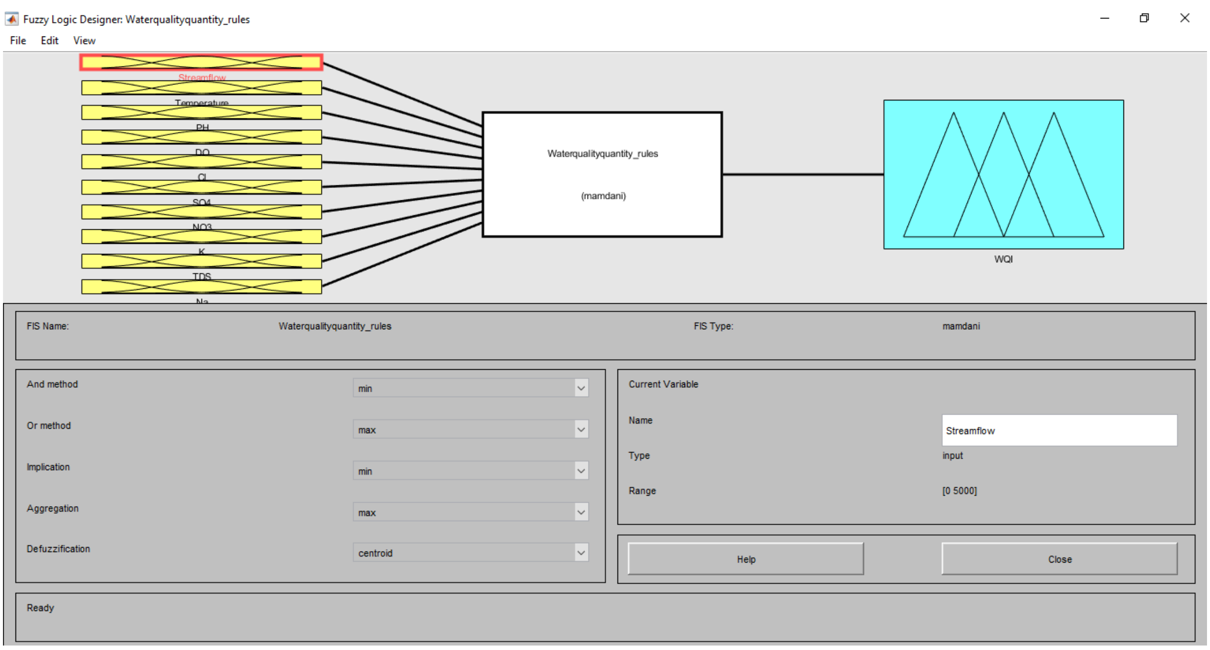


Figure 4.4 Mamdani fuzzy inference system with 10 inputs and 1 output

3.3.4 Hypothetical model simulation

Hypothetical modelling and simulation involve developing a model for various possible conditions for a particular study in cases where we have limited realistic data and resources. A hypothesis is necessarily a system of reasoning that involves those conditions whose reality is yet to be published (Gladun 1997). A few researches can be found in literature regarding hypothetical modelling in water quality management. Mujumdar and Sasikumar (2002) developed a fuzzy risk approach method for seasonal water quality management where they illustrated a fuzzy optimization model with a hypothetical river system. Rehana and Mujumdar (2011) studied the stream water quality response under various hypothetical climate change scenarios.

In our study, we have considered a case of moderately polluted stream as a hypothetical scenario for simulating monthly stream water quality. We have assumed a moderately polluted stream where the input water quality parameters are above their permissible limits by 25%. That means each of the quality parameter falls out of its permissible limit by 25%.

CHAPTER 4

STUDY AREA AND DATA USED

4.1 General

The study area in the present study is the Malaprabha River, which originates in the Western Ghats and flows through four districts of Karnataka State before joining the Krishna River. The Bennihalla River, which is one of the tributaries of Malaprabha, often floods the nearby villages and creates obstacles to the transport system of that region. Limited studies are reported with regard to the streamflow forecasting of the Malaprabha sub-basin at daily scale. Streamflow forecasting using soft computing methods and water quality modelling for this study area has not been previously attempted.

4.2 Description of the study area

The study region considered here is the watershed of Malaprabha River, a sub-basin of Krishna River basin. It is a sub-humid basin with the moisture index ranging between 0 and 20 (Thornthwaite 1948; Mudbhatkal et al. 2017). It has a catchment area of 11,549 km^{2}and extends between 15°30 N and 15°56 N latitudes and 74°12 E and 75°15 E longitudes (Figure 4.1).

The basin originates at a village called Kanakumbi inBelagavi district of Karnataka state in India at an altitude of 792.48m above mean sea level and flows for a length of 306 kms and joins the river Krishna at a place called Kudalasangama in Bagalkot district. The three sub-tributaries of Malaprabha River are Bennihalla, Hirehalla and Tuparihalla. The reservoir to the river Malaprabha at Saundatti known as Renuka Sagar reservoir supplies drinking water to Hubli-Dharwad city and facilitates the irrigation needs of Belagavi, Dharwad, Gadag and Bagalkot districts in Karnataka.

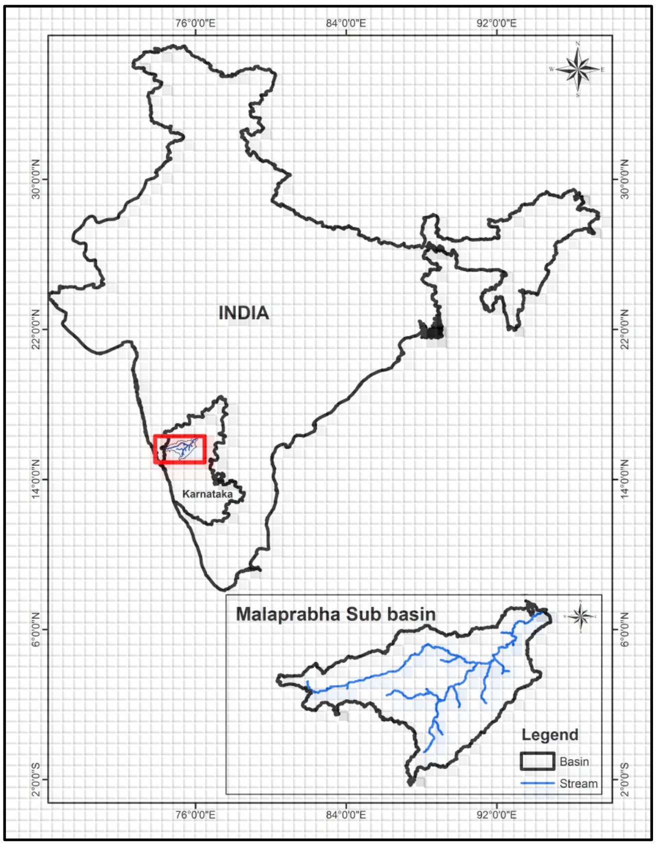


Figure 4.1 Basin map of the study region

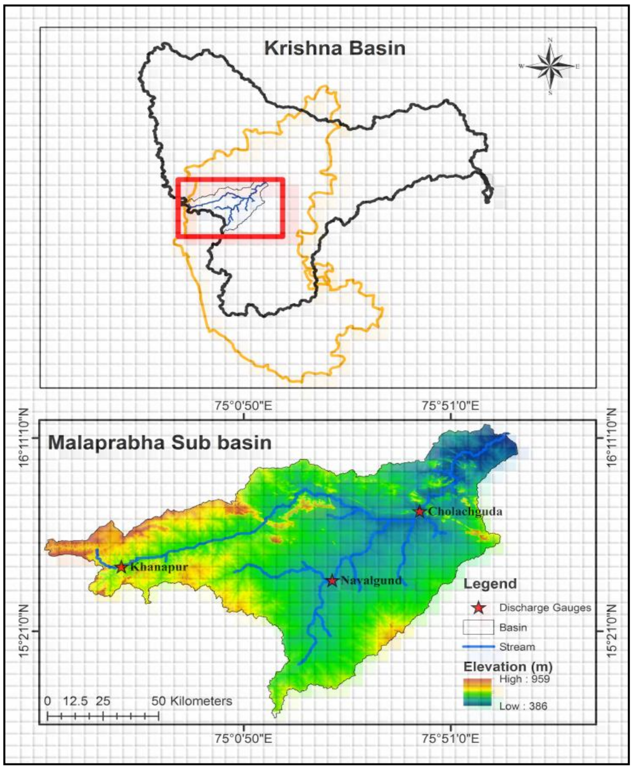


Figure 4.2 Location of discharge gauging stations and topography of study area

The gauging stations considered in this study are the Khanapur, Cholachguda and Navalgund which are situated in Belagavi, Bagalkot and Dharwad districts of Karnataka respectively (Figure 4.2).

4.3 Data collection

The details of data collection related to streamflow forecasting and water quality are discussed in the sub-sections below.

4.3.1 Streamflow data

The historical data related to the three stream gauging stations in the Malaprabha sub basin are collected and a brief description of streamflow data used in the study is given in Table 4.1. Khanapur station is at the upstream location of the study area and it is located in Belagavi district of Karnataka, whereas Navalgunda and Cholachguda stations are at the downstream of the basin and are located at Dharwad and Bagalkot districts of Karnataka state. The collected streamflow data was found to be in daily time scale and further the data was processed and checked for the missing data. The training of the models is carried out with 75% of the data and the testing is done for the remaining 25%.

Table 4.1Streamflow data description

Gauging station	**Period of record**	**Source of data**
Khanapur	1980-2018 (38 years)	Karnataka Water Resources Department, GOK
Cholachguda	1982-2015 (33 years)	India WRIS website
Navalgund	1990-2006 (16 years)	India WRIS website

4.3.2 Water quality data

The water quality data for Malaprabha river at the Khanapur station from June 2011 to May 2012 is presented in Table 4.2 (Sunkad 2013). A total of 9 parameters are measured for a period of one year on a monthly basis. These parameters along with the streamflow for the same period are considered for determining the Water quality index (WQI) using Fuzzy logic.

Table 4.2 Sampling data of different water quality parameters

Months	TDS (mg/l)	Temp (^{O}C)	pH	DO (mg/l)	Cl (mg/l)	Na (mg/l)	K (mg/l)	SO4 (mg/l)	NO3 (mg/l)
Jun 2011	170	22	7.62	7	18	8	1	9	3
Jul 2011	88	21	7.8	7.8	28	8.1	1.2	7	4.4
Aug2011	110	21	8	7.5	24	14.8	3	10	4
Sep 2011	120	22	7.6	8	26	16	1.2	20	6
Oct 2011	90	23	7.38	7.9	30	14.1	1.8	13	5
Nov2011	138	23	7.6	7.5	42	14.9	2.6	20	2
Dec 2011	140	25	7.9	7	40	16	1.9	20	3
Jan 2012	176	26	8	6	56	16.1	2	26	2
Feb 2012	260	28	8.2	7.1	85	20	2.1	32	4
Mar2012	265	29	7.12	5.3	86	16	1.9	32	1.87
Apr 2012	256	30	7.9	5.8	75	22	3	30	2
May2012	320	30	8.23	6	108	20	3.1	34.4	1.86

4.4 Summary

Malaprabha river supplies water to four districts of Karnataka state. It caters the needs of drinking water as well as irrigation. Streamflow forecasting for such a basin can facilitate various fields like food production, water demands, tourism and economy which are highly dependent on river flows. The streamflow data is measured at three gauging stations in Malaprabha sub-basin i. e., Khanapur, Cholachguda and Navalgund. On the other hand, limited water quality data is available for this particular basin. A lot of research has to be carried out related to water quality of Malaprabha river. For the present study, one year data related to various water quality parameters is adopted from literature.

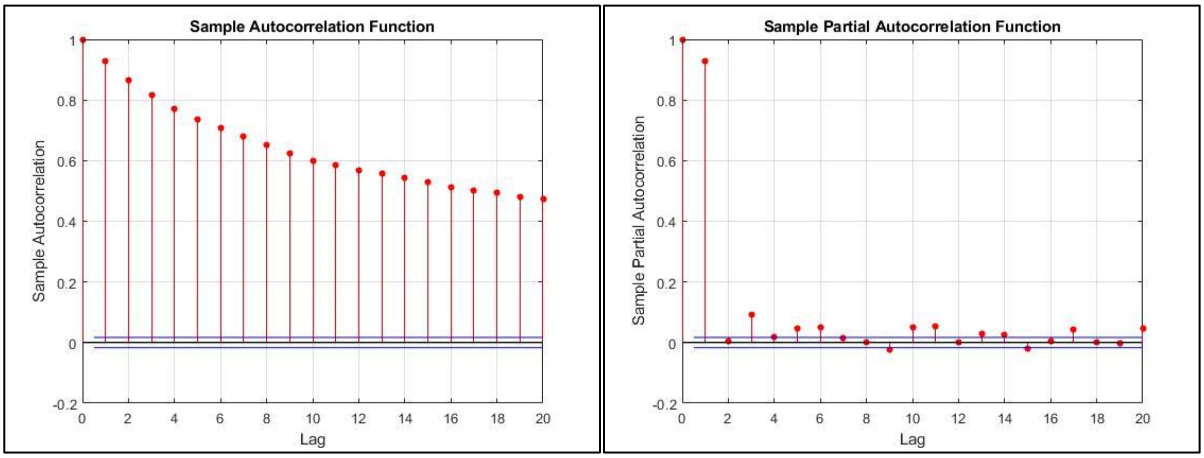
CHAPTER 5

RESULTS AND DISCUSSION

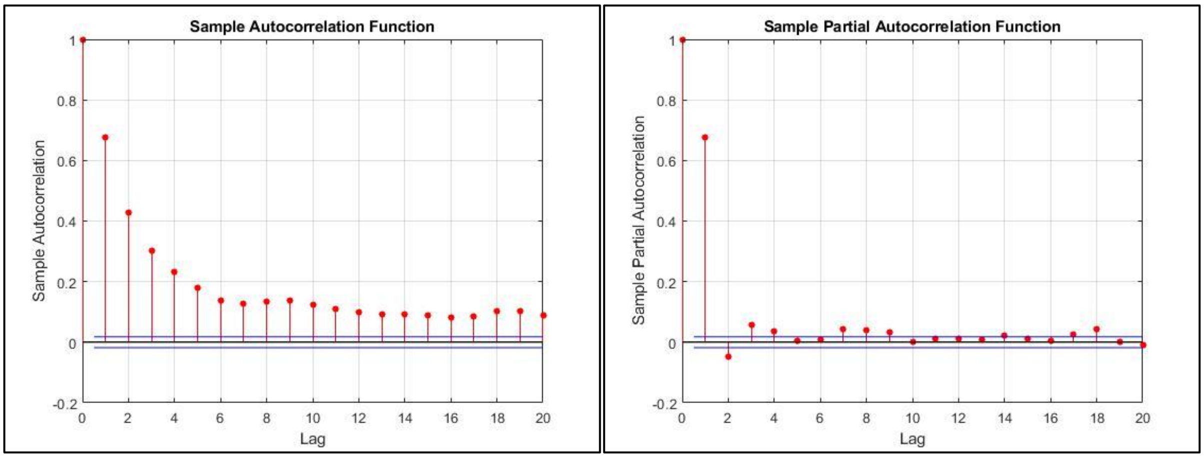
5.1 General

The number of lags (in days) in the time series for use in forecast models is identified based on the autocorrelation (ACF) and partial autocorrelation (PACF) functions. ACF is the correlation that exists between the series and the past and future values of the same series and PACF gives the partial correlation of a stationary time series with its own lagged values, regressed the values of the time series at all shorter lags. PACF considers both the actual correlation at a certain lag and the propagated correlation that is expected from previous lags (Hadi and Tombul, 2018)

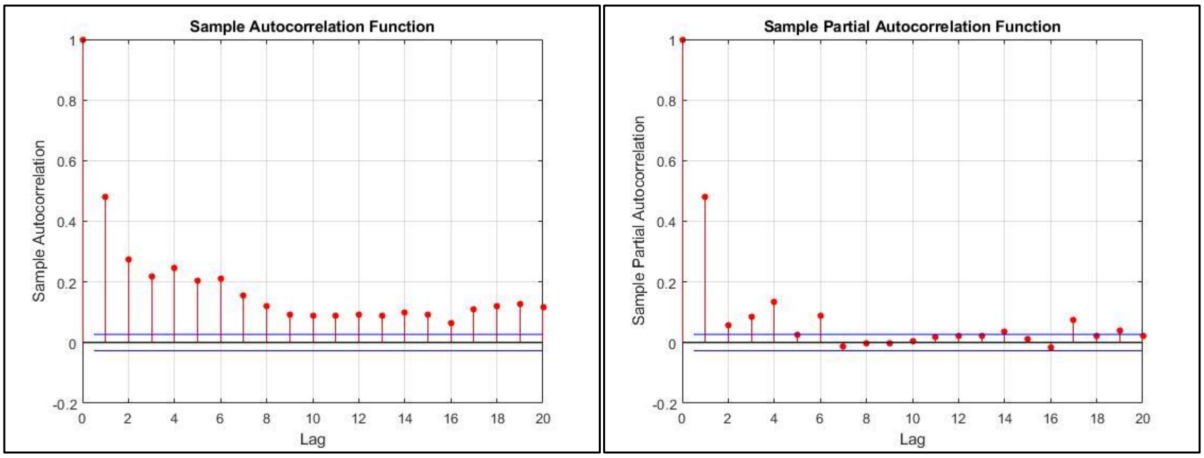
The number of lags is considered based on the ACF and PACF functions for each of the station. Figures 5.1 (a), (b) and (c) shows the ACF and PACF functions for the three stations respectively.



(a)



(b)



(c)

Figure 5.1 The autocorrelation and partial autocorrelation function plots for (a) Khanapur(b) Cholachguda (c) Navalgund

From Figure 5.1, the ACF function for all the three stations is gradually decreasing, which implies that the data is autoregressive in nature. But the correct order of lag is recognized by using PACF. In the PACF, the termination point of correlation is considered as the order of autoregressive lag (Machiwal and Jha 2012; Shumway and Stoffer 2010; Hadi and Tombul 2018). The ACF and PACF functions indicate that the series has a good correlation for 1-day lag only. After 1-day lag, the PACF shows a sudden drop which indicates that series is not autoregressive for 2-days and more lags. The PACF of all the three stations cut off at lag 1 and the correlation after lag 1 is low. Hence the models are considered only for lag 1.

The statistical characteristics of raw streamflow data are shown in Table 5.1, which indicate that the highest maximum streamflow and highest coefficient of variation

(C_v) for Cholachguda gauging station. In the present study, 75% of the data is considered for training the model and remaining 25% is used for testing. Except for Khanapur, other two stations have highest C_v for testing data set in comparison with the training data set.

Table 5.1 Statistical characteristics of streamflow data without pre-processing

Gauging Station	Minimum (m^3/s)		Maximum (m^3/s)		Standard deviation (m^3/s)		Coefficient of Variation		Skewness	
	Train Data	Test Data	Train Data	Test Data	Train Data	Test Data	Train Data	Test Data	Train Data	Test Data
Khanapur	0.00	0.00	745.5	655.11	69.83	71.84	225.96	196.66	1.32	1.53
Cholachguda	0.00	0.00	1814	2751.89	80.22	107.6	288.76	406.60	0.64	0.46
Navalgund	0.00	0.00	407.8	338.07	25.18	22.41	300.75	401.31	0.66	0.74

5.2 Data pre-processing and wavelet analysis

Table 5.2 shows the various mother wavelets and their types used in the study. Each of the type of mother wavelet is decomposed up to level 5. A detailed description of wavelet theory and applications can be found in Graps 1995.

Table 5.2 Various mother wavelets used in the study

Daubechies	Symlets	Haar or db1	Coiflets
db2 db3 db4 db5 db6 db7	sym1 sym2 sym3 sym4 sym5 sym6		coif1 coif2 coif3 coif4 coif5

5.2.1 Choice of mother wavelet function

Figures 5.2, 5.3 and 5.4 show the correlation of various mother wavelets with 5 decomposition levels for SVM model. Highest correlated mother wavelet and its corresponding decomposition level are chosen as best suitable mother wavelet for that particular station. Similar method is used for ANFIS and Fuzzy models.

It is evident from Figures 5.2, 5.3 and 5.4 that the choice of mother wavelet is station or region sensitive (also, sensitive to the data used). The Mother Wavelets are sensitive to the characteristics of location of stream gauging stations as well as the length of the data. The results of the study indicated that the best suitable mother wavelet and decomposition level for each stream gauging station is different. The best suitable mother wavelet functions for Khanapur, Cholachguda and Navalgund gauging stations are Haar and coif at level 2, db6 at level 2, and Haar at level 2 respectively.

Similar analysis is carried out for ANFIS and Fuzzy models for the choice of best suitable mother wavelet and the decomposition level. The results for ANFIS model show that db2 at level 2, db6 at level 3 and haar at level 3 are the best suitable mother wavelets for Khanapur, Cholachguda and Navalgund stations respectively. Whereas, the fuzzy model results show that coif at level 2, db2 at level 2 and haar at level 2 are the best suitable mother wavelets for Khanapur, Cholachguda and Navalgund stations respectively.

From wavelet analysis, we can infer that the choice of mother wavelet is sensitive to length of the data, region of study and flow rate. It is evident from Tables 5.4, 5.6 and 5.8 that, for a particular station data, mother wavelet remains same irrespective of the models used. However, the decomposition levels vary for each station and each model.

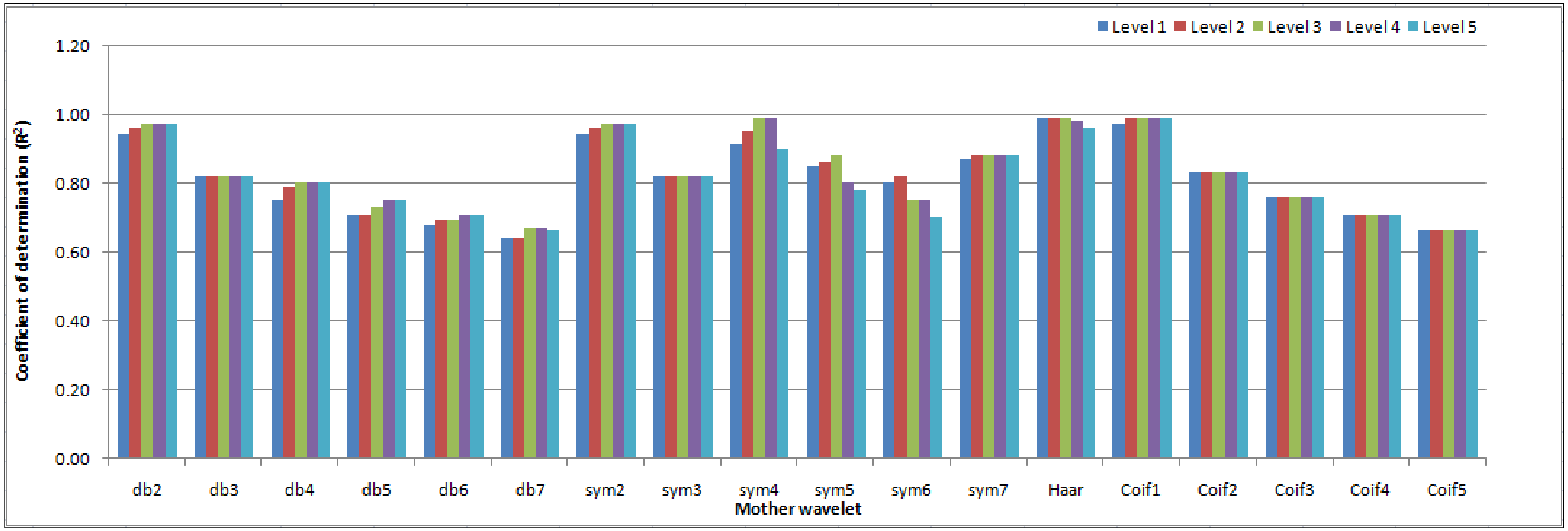


Figure 5.2: Performance of various mother wavelets with level 5 decomposition using SVM for the test phase of Khanapur gauging station

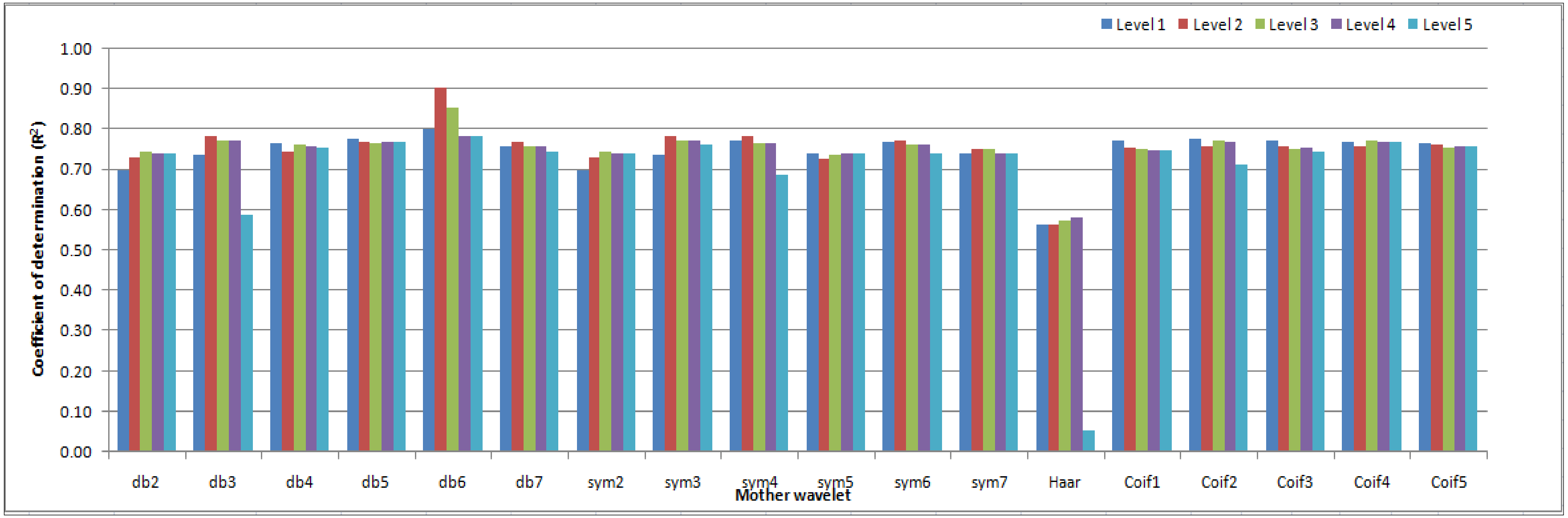


Figure 5.3: Performance of various mother wavelets with level 5 decomposition using SVM for the test phase of Cholachguda gauging station

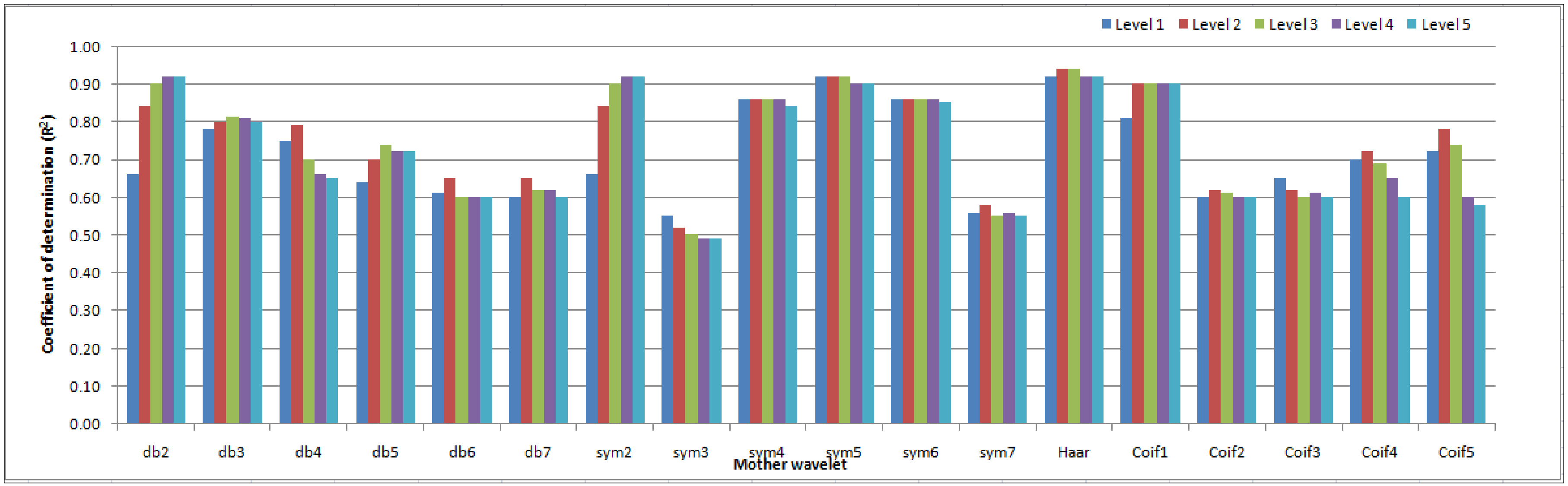


Figure 5.4: Performance of various mother wavelets with level 5 decomposition using SVM for the test phase of Navalgund gauging station

5.3 Streamflow forecasting using various soft computing methods

The results of single soft computing models and corresponding wavelet couple soft computing models are presented below.

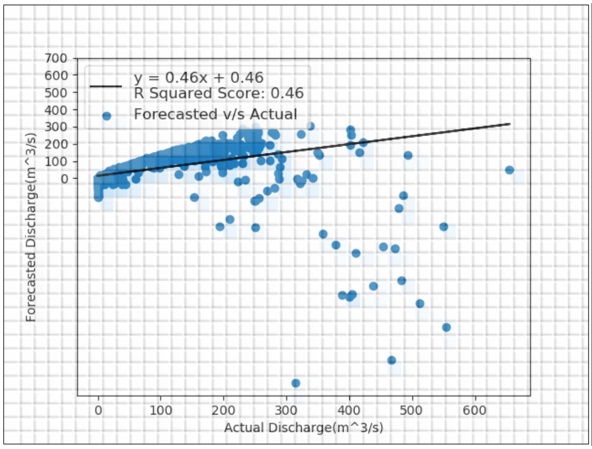
5.3.1 Results of SVM and WT-SVM models

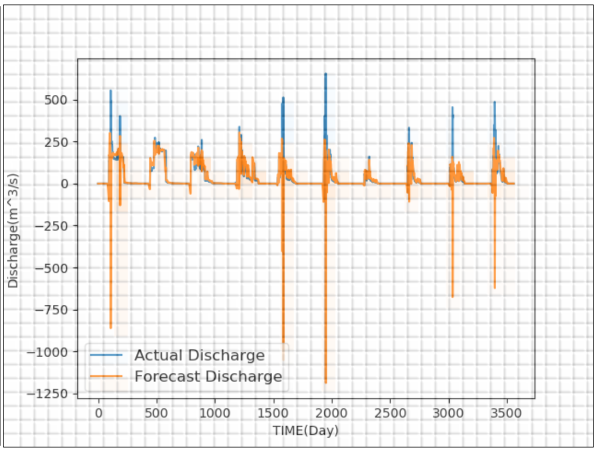
The hyper-parameters of SVM model are optimized using Sequential minimal optimization algorithm in MATLAB interface. Table 5.3 shows the performance of single SVM model for the all the three stations.

Table 5.3 Results of single SVM model for 1-day, 3-days and 5-days ahead forecasts

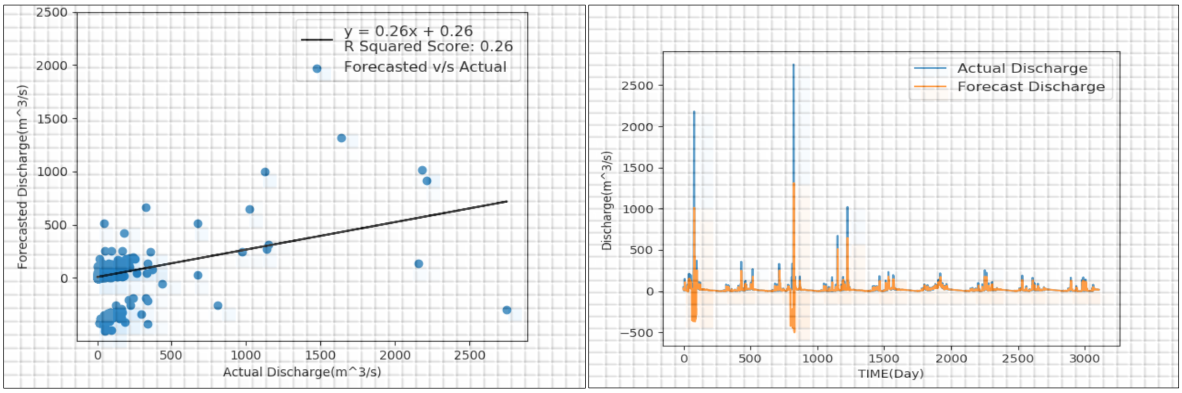
Performance Index	Station								
	Khanapur			Cholachguda			Navalgund		
	1 day	3 days	5 days	1 day	3 days	5 days	1 day	3 days	5 days
R^2	0.914	0.62	0.46	0.80	0.59	0.26	0.88	0.65	0.39
RMSE(m^3/s)	21.382	45.65	60.73	58.88	75.14	148.79	8.40	15.32	25.23
NSE	0.90	0.60	0.40	0.77	0.62	0.25	0.79	0.61	0.35

The results for single SVM model show significant decrease in R^2 and NSE values and an increase in RMSE from 1-day to 3-days ahead and from 3-days to 5-days ahead forecast values. The forecasting accuracy decreases with the increase in lead time.

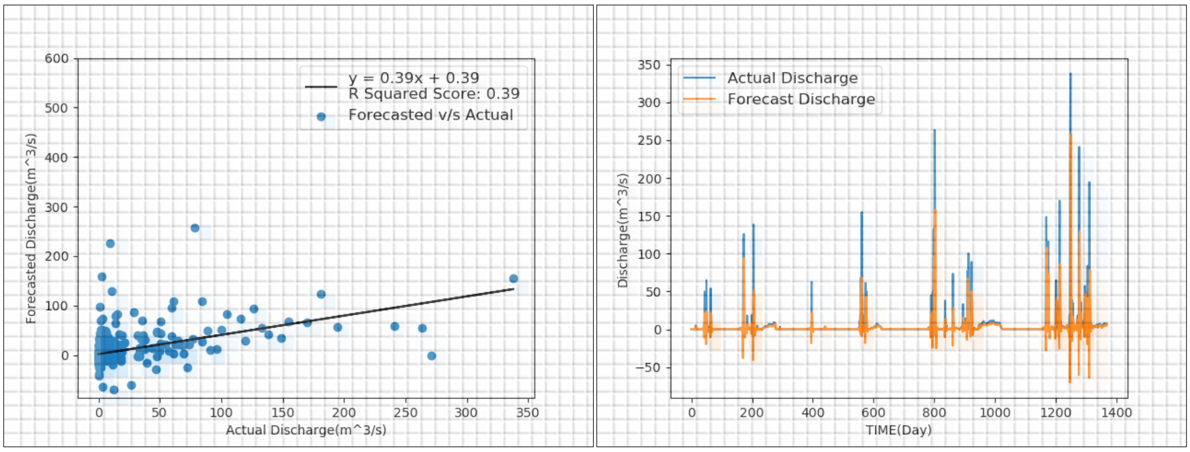




(a)



(b)

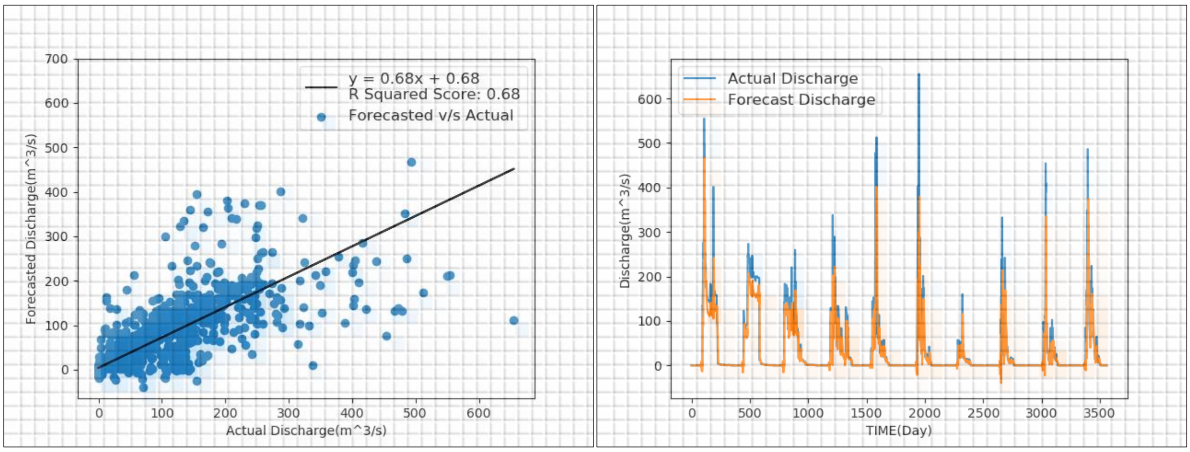


(c)

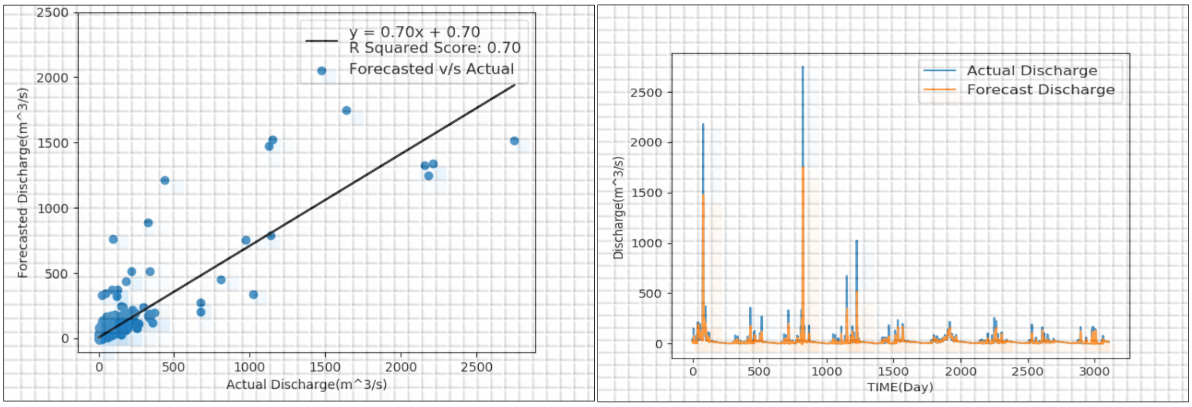
Figure 5.5 Scatter plots of single SVM model for 5-days ahead forecast for test period of (a) Khanapur (b) Cholachguda (c) Navalgund gauging stations

Table 5.4 Performance of Wavelet coupled SVM model for the three stations with their best suitable mother wavelets for 1-day, 3-days and 5-days ahead forecasts.

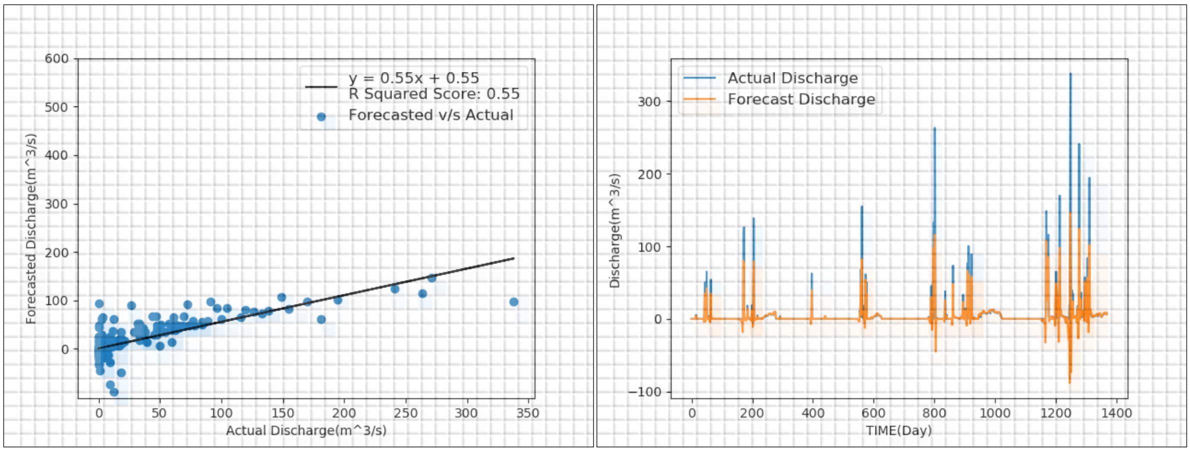
Performance index ↓	**Station →**	**Khanapur**			**Cholachguda**			**Navalgund**		
	Mother wavelets →	**Haar and Coif1 at level 2**			**Daubechies (db6 at level 2)**			**Haar at level 2**		
		1-day	**3-days**	**5-days**	**1-day**	**3-days**	**5-days**	**1-day**	**3-days**	**5-days**
R^2		**0.99**	**0.78**	**0.68**	**0.90**	**0.71**	**0.70**	**0.94**	**0.79**	**0.55**
RMSE(m^3/s)		**18.25**	**38.36**	**53.26**	**50.82**	**62.36**	**79.42**	**6.25**	**13.52**	**22.58**
NSE		**0.98**	**0.75**	**0.66**	**0.88**	**0.70**	**0.68**	**0.93**	**0.76**	**0.52**



(a)



(b)



(c)

Figure 5.6 Scatter plots of wavelet coupled SVM model for 5-days ahead forecast for test period of (a) Khanapur (b) Cholachguda (c) Navalgund gauging stations

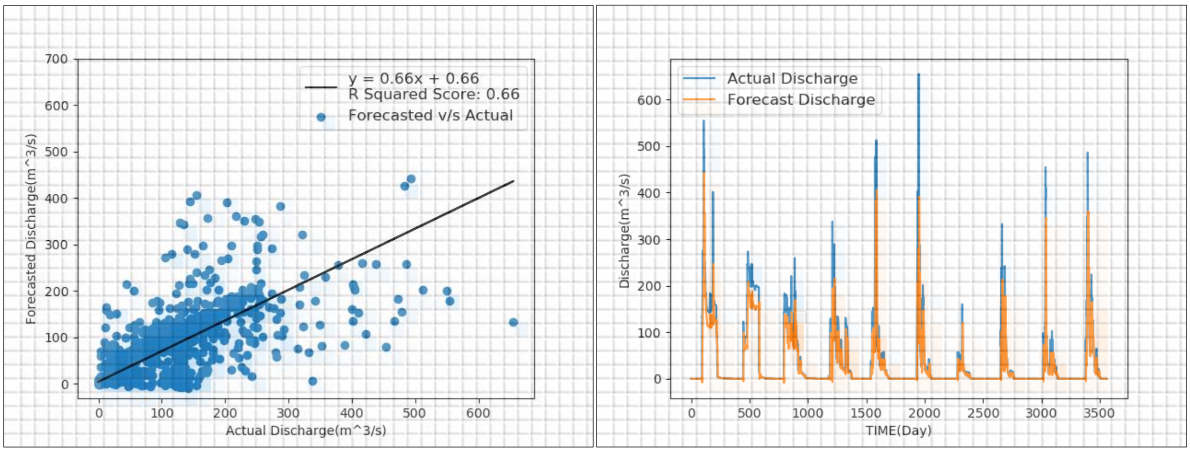
As seen in Table 5.4, the performance of SVM is enhanced for all the three stations when it is integrated with wavelet. There is a subsequent improvement in the results of 5-day lead time forecasts, which shows that usage of pre-processing technique like wavelet that increase the accuracy of higher lead time forecast values efficiently. The result is in agreement with Kisi and Cimen (2011) who used wavelet-SVM model to forecast monthly flow series. It is found that the conjunction model with the value of R^2 = 0.768 is superior to the most accurate SVM model with R^2 = 0.525. Whereas, Hadi and Tombul (2018) used wavelet transform as a data pre-processing tool to forecast 7 days ahead streamflow using data driven models and found that artificial neural network (ANN) model is superior to adaptive neuro fuzzy inference system (ANFIS) and SVM models. As observed from literature survey, comparison of various mother wavelets for daily streamflow forecasting application has not been reported in the literature so far.

5.3.2 Results of single ANFIS model and Wavelet coupled ANFIS model (WT-ANFIS)

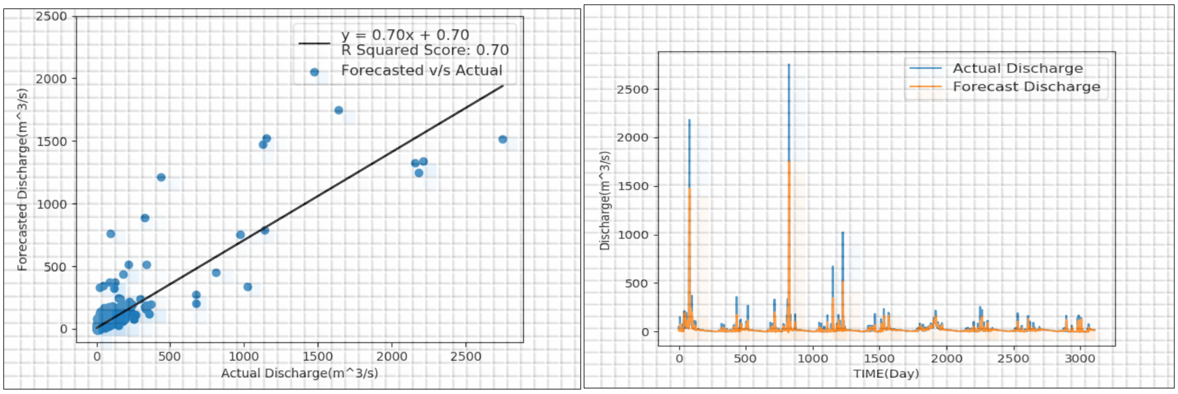
The adaptive neuro fuzzy inference system (ANFIS) model itself is a hybrid model of Artificial Neural Network (ANN) and Fuzzy logic. The performance of ANFIS for 1-day, 3-days and 5-days ahead stream flow forecasting is as shown in Table 5.5.

Table 5.5 Results of ANFIS model for 1-day, 3-days and 5-days ahead forecasts

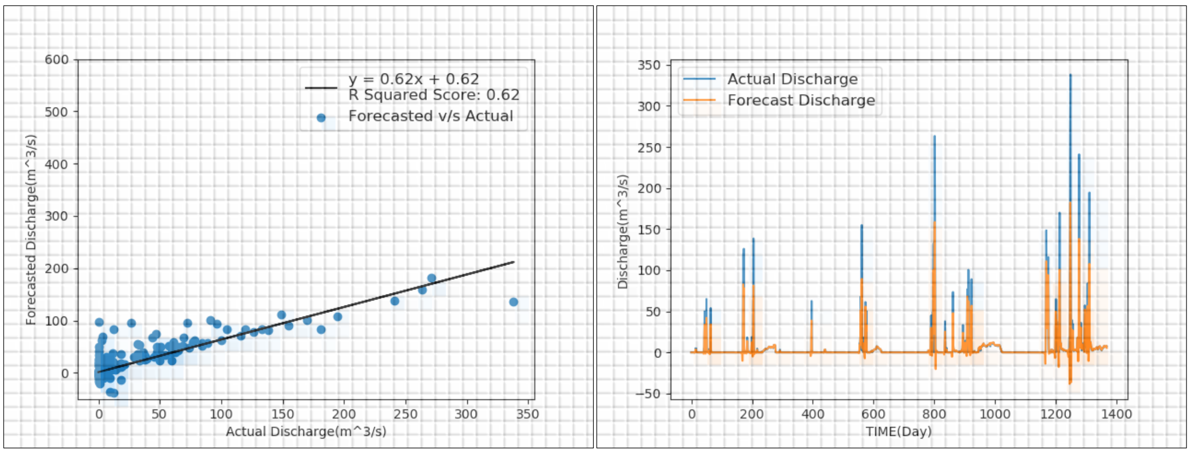
Performance Index	Station								
	Khanapur			Cholachguda			Navalgund		
	1 day	3 days	5 days	1 day	3 days	5 days	1 day	3 days	5 days
R^2	0.92	0.78	0.66	0.82	0.70	0.70	0.86	0.62	0.62
RMSE(m^3/s)	21.45	35.28	42.56	58.88	70.54	70.54	8.28	14.6	18.39
NSE	0.91	0.75	0.65	0.80	0.69	0.69	0.87	0.61	0.61



(a)



(b)



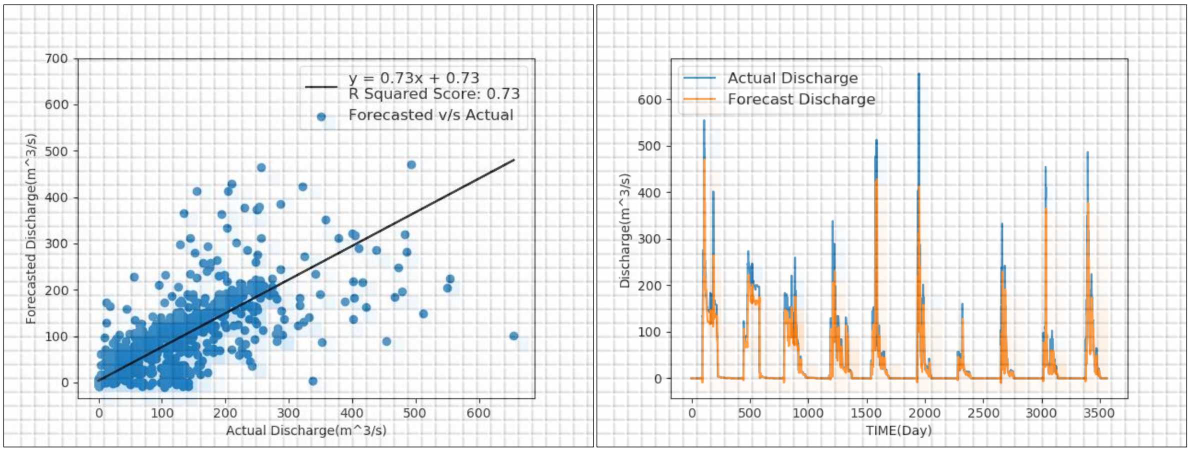
(c)

Figure 5.7 Scatter plots of single ANFIS model for 5-days ahead forecast for test period of (a) Khanapur (b) Cholachguda (c) Navalgund gauging stations

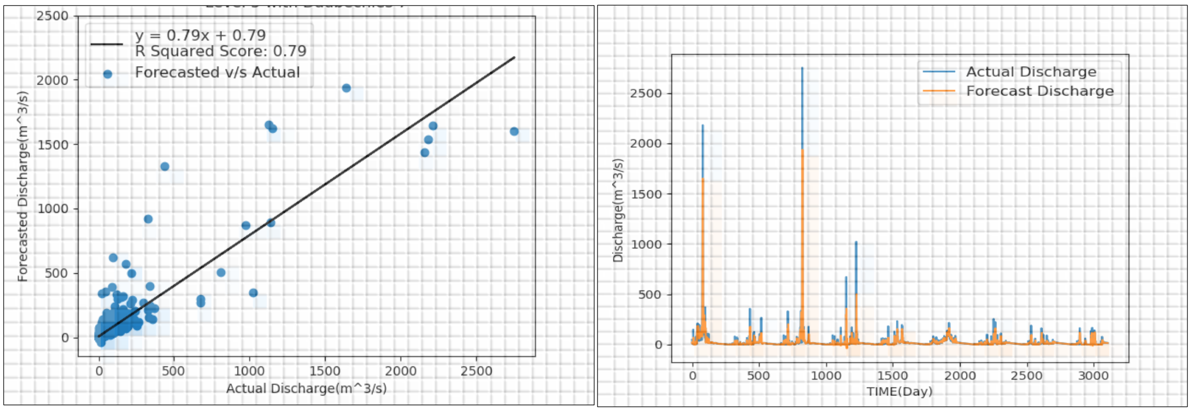
It can be observed from Table 5.5 that ANFIS model's performance is at its best for 1-day ahead forecasting. The performance of the model degraded as the lead time is increased to 5 days as well as the model showed underestimation of peaks for all the lead times (Fig. 5.7). That means ANFIS model failed to capture peaks. Hence this model can be suitable only for medium and low flow rates.

Table 5.6 Performance of Wavelet coupled ANFIS model for the three stations with their best suitable mother wavelets for 1-day, 3-days and 5-days ahead forecasts.

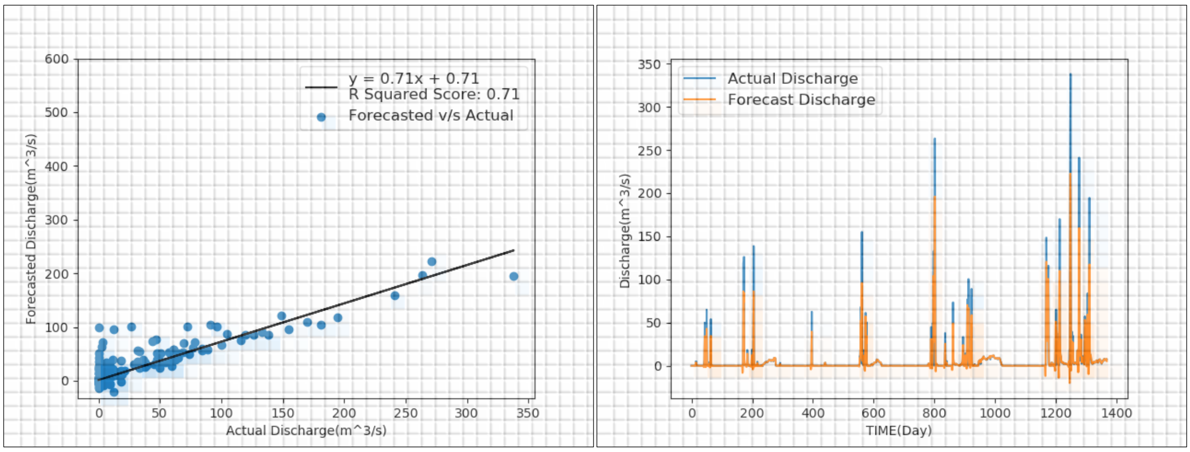
Performance index ↓	**Station →**	**Khanapur**			**Cholachguda**			**Navalgund**		
	Mother wavelets →	**db2 at level 2**			**Daubechies (db6 at level 3)**			**Haar at level 3**		
		1-day	**3-days**	**5-days**	**1-day**	**3-days**	**5-days**	**1-day**	**3-days**	**5-days**
R^2		**0.99**	**0.81**	**0.73**	**0.94**	**0.80**	**0.79**	**0.98**	**0.83**	**0.71**
RMSE(m^3/s)		**18.30**	**35.42**	**50.46**	**40.62**	**60.21**	**62.36**	**5.26**	**11.56**	**18.39**
NSE		**0.98**	**0.80**	**0.71**	**0.92**	**0.80**	**0.79**	**0.97**	**0.82**	**0.70**



(a)



(b)



(c)

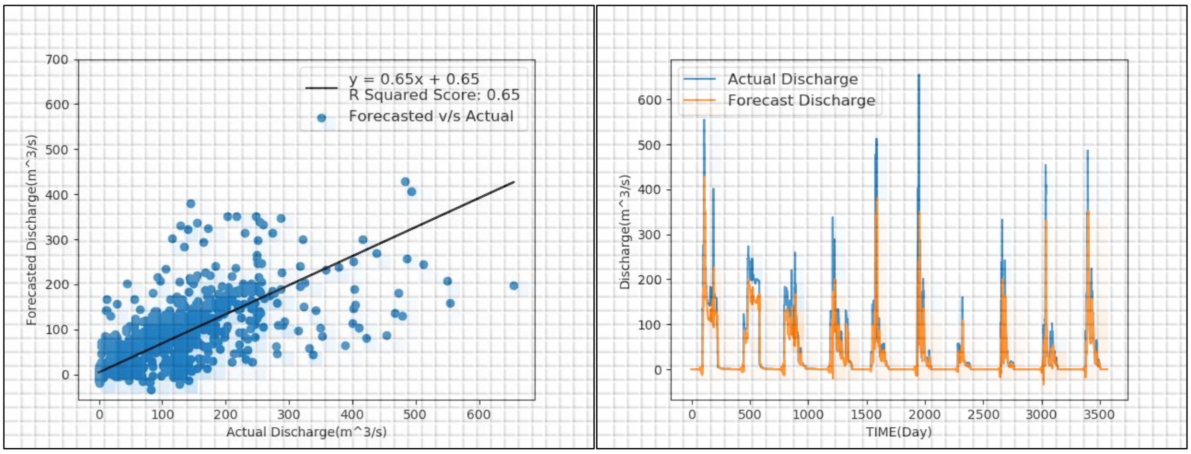
Figure 5.8 Scatter plots of wavelet coupled ANFIS model for 5-days ahead forecast for test period of (a) Khanapur (b) Cholachguda (c) Navalgund gauging stations

From Table 5.6 and Fig. 5.8, it is evident that the performance of forecasting model, that is, ANFIS improved substantially after it was coupled with wavelet method. In Fig. 5.7 and 5.8, it is observed that there are some negative streamflow values for 5-days ahead forecasting. Though there is improvement in the performance, the model underestimated the peaks. The negative streamflow values can be attributed to the uncertainties associated with the data and modelling.

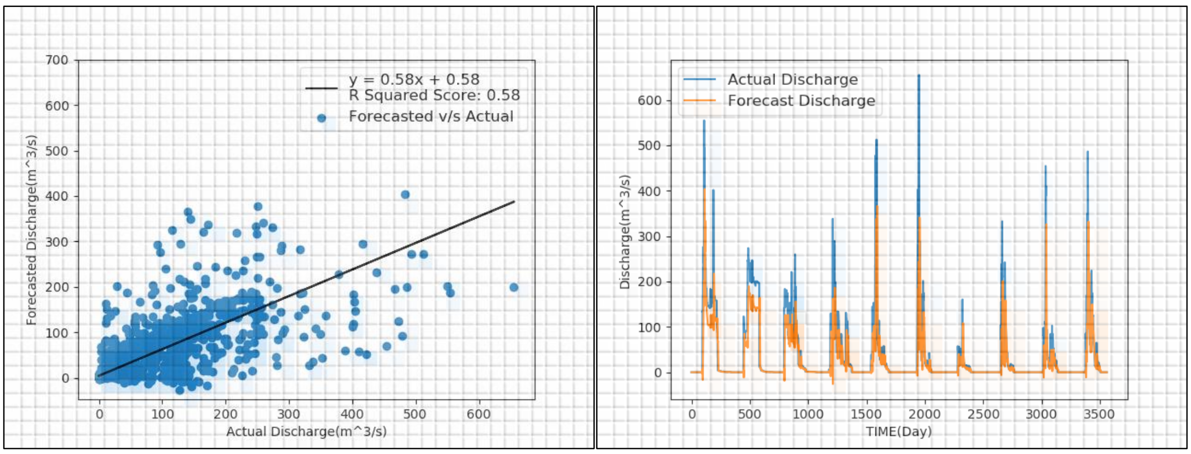
5.3.3 Results of single Fuzzy model and Wavelet coupled Fuzzy model (WT-Fuzzy)

Table 5.7Results of single Fuzzy model for 1-day, 3-days and 5-days ahead forecasts

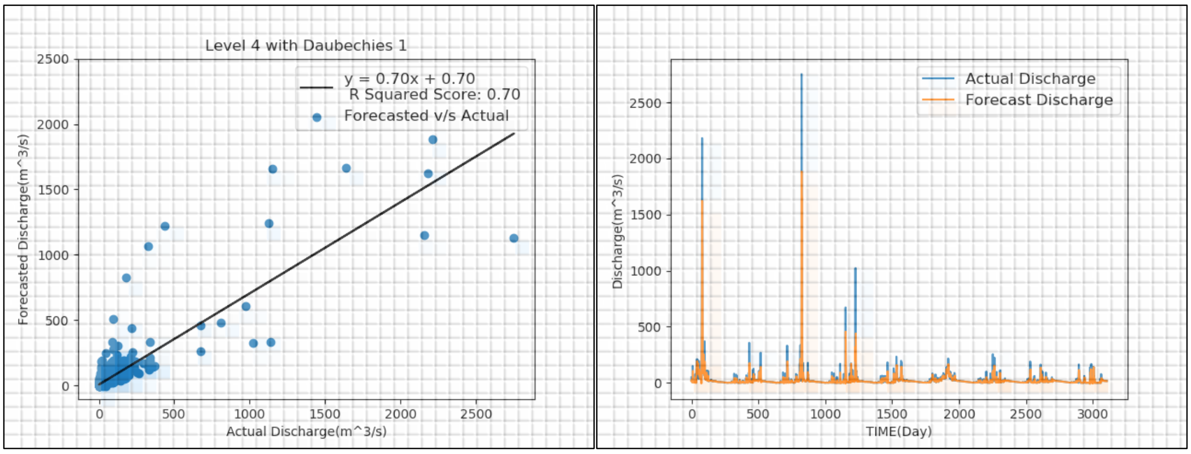
Performance Index	Station								
	Khanapur			Cholachguda			Navalgund		
	1 day	3 days	5 days	1 day	3 days	5 days	1 day	3 days	5 days
R^2	0.94	0.68	0.65	0.86	0.63	0.58	0.81	0.76	0.70
RMSE(m^3/s)	19.94	42.80	43.40	52.64	68.74	76.23	7.23	11.52	14.60
NSE	0.92	0.65	0.62	0.85	0.62	0.60	0.79	0.74	0.69



(a)



(b)



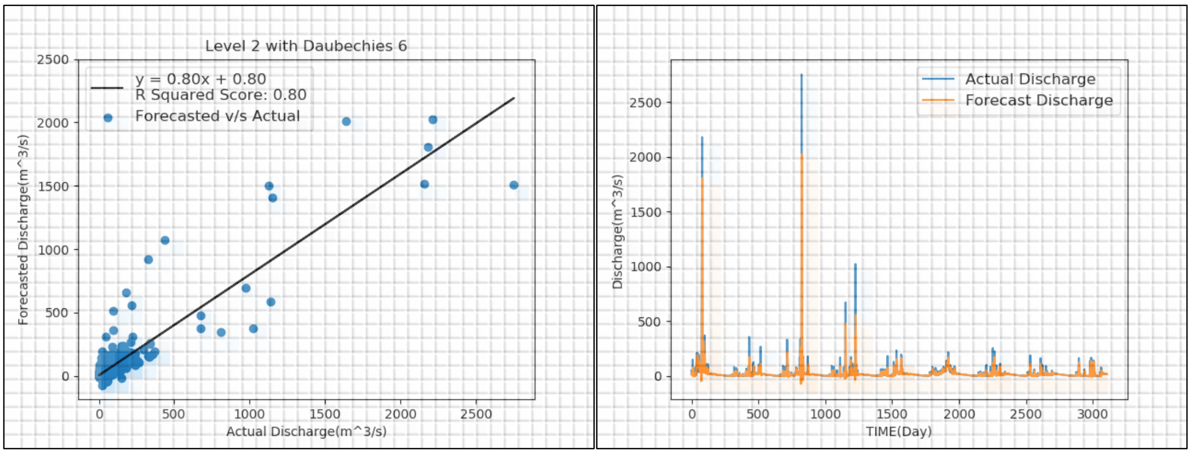
(c)

Figure 5.9 Scatter plots of single Fuzzy model for 5-days ahead forecast for test period of (a) Khanapur (b) Cholachguda (c) Navalgund gauging stations

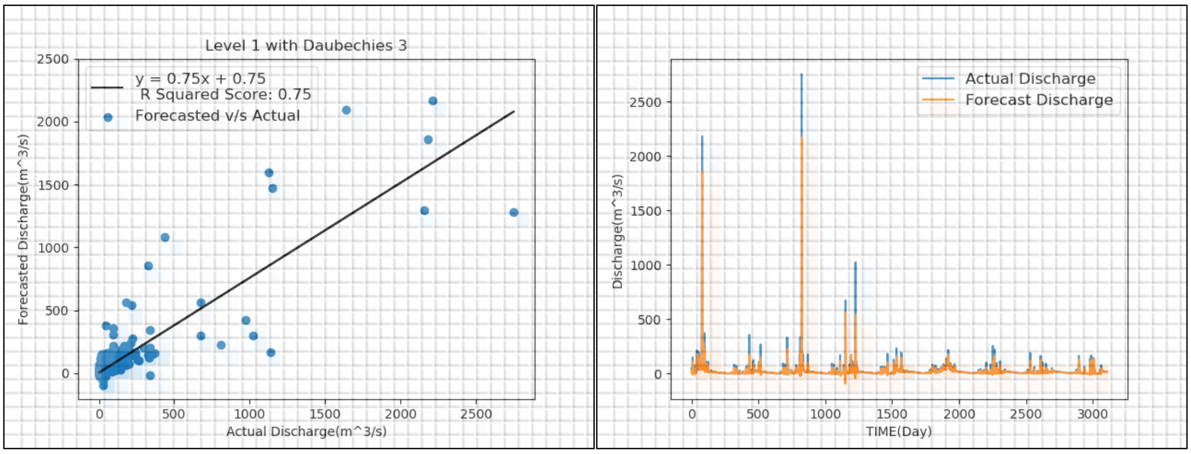
The results of single fuzzy model are presented in Table 5.7 and Fig. 5.9. It can be inferred that fuzzy model degraded for longer lead times. The model certainly underestimated the peaks. The negative values of streamflow for Khanapur and Cholachguda stations can be attributed to the uncertainly of data and model itself.

Table 5.8 Performance of Wavelet coupled fuzzy model for the three stations with their best suitable mother wavelets for 1-day, 3-days and 5-days ahead forecasts.

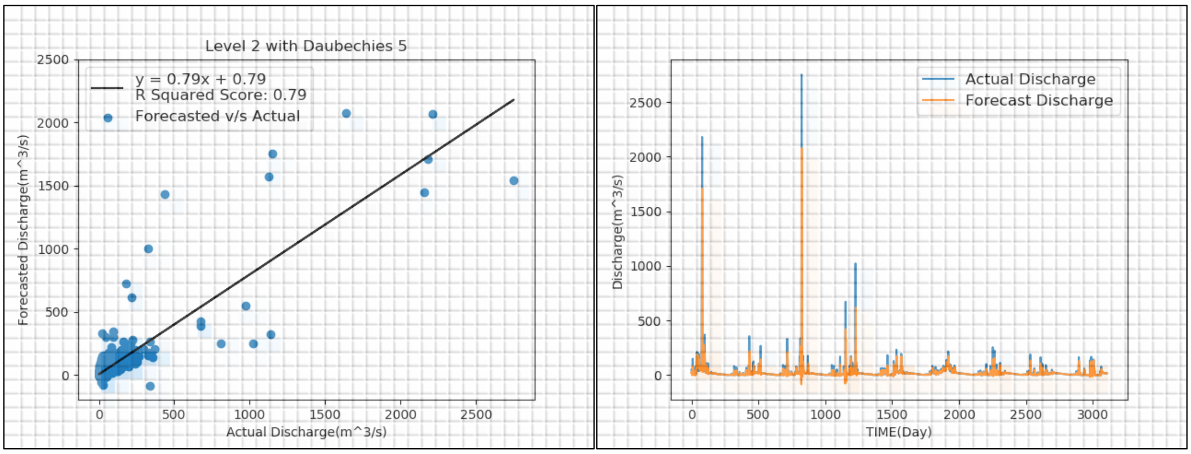
Performance index ↓	Station →	Khanapur			Cholachguda			Navalgund		
	Mother wavelets →	Coif1 at level 2			Daubechies (db2 at level 2)			Haar at level 3		
		1-day	3-days	5-days	1-day	3-days	5-days	1-day	3-days	5-days
R^2		0.99	0.94	0.80	0.98	0.84	0.75	0.95	0.80	0.79
RMSE(m^3/s)		19.56	20.29	40.14	38.45	60.16	60.28	6.68	12.34	13.51
NSE		0.98	0.93	0.78	0.97	0.83	0.73	0.93	0.78	0.76



(a)



(b)



(c)

Figure 5.10 Scatter plots of wavelet coupled Fuzzy model for 5-days ahead forecast for test period of (a) Khanapur (b) Cholachguda (c) Navalgund gauging stations

The results of wavelet coupled fuzzy model are presented in Table 5.8 and Fig. 5.10, which indicate that the performance of single fuzzy model can be significantly improved by integrating wavelet method for data pre-processing. The hybrid model showed some negative discharges which may be due to various uncertainties associated with modelling.

Negative discharges may be attributed to the model's inefficiency and uncertainties of modelling. In some cases, the stream gauge readings can go below a marked "zero" value (during dry season) which may be read as negative values in modeling but in reality, those are lowest flows in the stream.

5.4 Comparison of hybrid wavelet models

Here in this study, we have considered the percentage increase or decrease in the correlation coefficient as the measure for comparison of hybrid wavelet models. Tables 5.9, 5.10 and 5.11 show the comparison of single models and wavelet hybrid models.

Table 5.9 Summary of results and percentage increase in SVM model performance with wavelet

Station	**R^2value**								
	1-day ahead		**% increase**	**3-days ahead**		**% increase**	**5-days ahead**		**% increase**
	SVM	**WT-SVM**		**SVM**	**WT-SVM**		**SVM**	**WT-SVM**	
Khanapur	**0.914**	**099**	**8.32**	**0.66**	**0.73**	**10.61**	**0.46**	**0.68**	**47.83**
Cholachguda	**0.80**	**0.90**	**12.5**	**0.70**	**0.79**	**12.85**	**0.26**	**0.70**	**169.23**
Navalgund	**0.88**	**0.94**	**6.82**	**0.62**	**0.71**	**14.52**	**0.39**	**0.55**	**41.03**

Table 5.10 Summary of results and percentage increase in ANFIS model performance with wavelet

Station	R^2value								
	1-day ahead		% increase	3-days ahead		% increase	5-days ahead		% increase
	ANFIS	WT-ANFIS		ANFIS	WT-ANFIS		ANFIS	WT-ANFIS	
Khanapur	**0.92**	**0.99**	**7.61**	**0.78**	**0.81**	**3.85**	**0.66**	**0.73**	**10.61**
Cholachguda	**0.82**	**0.94**	**14.63**	**0.70**	**0.80**	**14.23**	**0.70**	**0.79**	**12.86**
Navalgund	**0.98**	**0.98**	**13.95**	**0.62**	**0.83**	**33.87**	**0.62**	**0.71**	**14.52**

Table 5.11 Summary of results and percentage increase in Fuzzy model performance with wavelet

Station	R^2value								
	1-day ahead		% increase	3-days ahead		% increase	5-days ahead		% increase
	Fuzzy	WT-Fuzzy		Fuzzy	WT-Fuzzy		Fuzzy	WT-Fuzzy	
Khanapur	**0.94**	**0.99**	**5.32**	**0.68**	**0.94**	**38.24**	**0.65**	**0.80**	**23.08**
Cholachguda	**0.86**	**0.98**	**13.95**	**0.63**	**0.84**	**33.33**	**0.58**	**0.75**	**29.31**
Navalgund	**0.81**	**0.95**	**17.28**	**0.76**	**0.80**	**5.26**	**0.70**	**0.79**	**12.86**

It is evident from the above Tables 5.9, 5.10, 5.11 that, ANFIS and Fuzzy models outperformed SVM model. Fuzzy model performed better than ANFIS. When coupled with wavelet, all the model performance has been enhanced for all the three models. This shows the importance of data pre-processing in hydrological modelling. The results of WT-ANFIS and WT-Fuzzy models are very close in terms of performance. Both the models performed better than WT-SVM. However, WT-Fuzzy outperformed the other two models.

5.5 Stream Water quality-quantity model

Fuzzy logic water quality is developed to determine the water quality index for Malaprapha river at Khanapur gauging station. The river water quality data for Malaprabha river at the Khanapur station is measured from June 2011 to May 2012 (Sunkad 2013). A total of 9 parameters are measured for a period of one year on a monthly basis. These parameters along with the streamflow for the same period are considered for determining the Water quality index (WQI) using Fuzzy logic. The details of input and output parameters are presented in Table 5.12.

Table 5.12 The input and output parameters for the proposed fuzzy model

Inputs	Output
pH	Water quality index
Electrical conductivity (Ec) (µMhos/cm)	
Chloride (Cl) (mg/l)	
Sodium (Na) (mg/l)	
Potassium (K) (mg/l)	
Calcium (Ca) (mg/l)	
Magnesium (Mg) (mg/l)	
Nitrate (No3) (mg/l)	
Sodium adsorption ratio (SAR)	
Streamflow (m^3/s)	

5.5.1 Calculation of WQI

Water quality models can be effective tools to simulate and predict pollutant transport in water environment, which can contribute to saving the cost of labours and materials for a large number of chemical experiments to some degree. Moreover, it is inaccessible for on-site experiments in some cases due to special environmental pollution issues. Therefore, water quality models become an important tool to identify water environmental pollution and the final fate and behaviours of pollutants in water environment (Wang et al. 2013).

The water quality of the Malaprabha River at Khanapur station was evaluated using Water Quality Index (WQI). The standards for various water quality parameters used in the study are taken from BIS 2012 and FAO 1985. The calculation of relative weights of each parameter is illustrated in Table 5.13. The WQI for the given water quality parameters is calculated for each month as shown in Table 5.14.

Table 5.13 Calculation of relative weights of each parameter

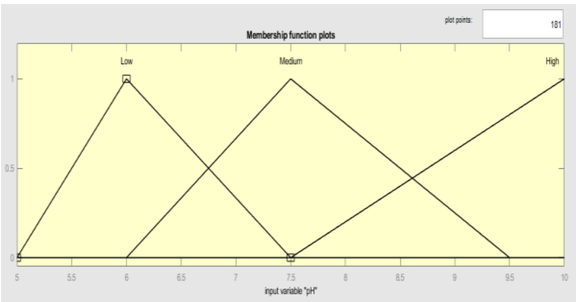
Inputs	Standards (BIS and FAO)	Weight	Relative weight
P^H	6 to 8.5	1	0.0909
Electrical conductivity (Ec) (µMhos/cm)	1000	1	0.0909
Chloride (Cl) (mg/l)	600	1	0.0909
Sodium (Na) (mg/l)	920	1	0.0909
Potassium (K) (mg/l)	2	3	0.2727
Calcium (Ca) (mg/l)	400	1	0.0909
Magnesium (Mg) (mg/l)	60	1	0.0909
Nitrate (No3) (mg/l)	45	1	0.0909
Sodium adsorption ratio (SAR)	26	1	0.0909

Table 5.14 Calculation of WQI

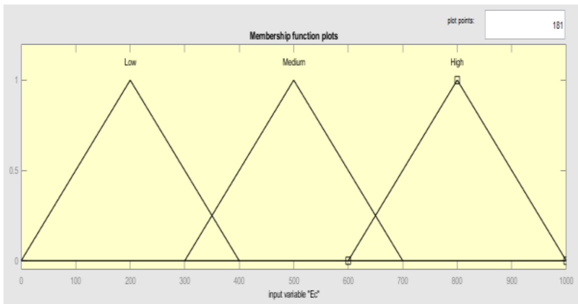
Months	**PH**	qi	SIi	**Ec**	qi	SIi	**Cl**	qi	SIi	**Na**	qi	SIi	**K**	qi	SIi	**Ca**	qi	SIi	**Mg**	qi	SIi	**NO3**	qi	SIi	**SAR**	qi	SIi	**WQI**	**Class**
jun	7.62	89.6	8.15	110	11	1	18	3	0.3	8	0.87	0.08	1	50	13.6	16	4	0.4	3.2	5.25	0.48	3	6.67	0.61	2.59	9.94	0.9	25.5	Execellent
jul	7.8	91.8	8.34	140	14	1.3	28	4.7	0.4	8.1	0.88	0.08	1.2	60	16.4	13	3.2	0.3	3.5	5.87	0.53	4.4	9.78	0.89	2.85	11	1	29.2	Execellent
aug	8	94.1	8.56	160	16	1.5	24	4	0.4	15	1.61	0.15	3	150	40.9	11	2.8	0.2	2.6	4.25	0.39	4	8.89	0.81	5.69	21.9	1.99	54.9	Good
sep	7.6	89.4	8.13	180	18	1.6	26	4.3	0.4	16	1.74	0.16	1.2	60	16.4	12	3	0.3	2.4	4.05	0.37	6	13.3	1.21	5.96	22.9	2.08	30.6	Execellent
oct	7.38	86.8	7.89	130	13	1.2	30	5	0.5	14	1.53	0.14	1.8	90	24.5	8.4	2.1	0.2	1.2	2.02	0.18	5	11.1	1.01	6.43	24.7	2.25	37.8	Execellent
nov	7.6	89.4	8.13	210	21	1.9	42	7	0.6	15	1.62	0.15	2.6	130	35.5	15	3.8	0.3	2.6	4.38	0.4	2	4.44	0.4	5.02	19.3	1.75	49.2	Execellent
dec	7.9	92.9	8.45	210	21	1.9	40	6.7	0.6	16	1.74	0.16	1.9	95	25.9	14	3.5	0.3	4.1	6.88	0.63	3	6.67	0.61	5.31	20.4	1.86	40.4	Execellent
jan	8	94.1	8.56	268	27	2.4	56	9.3	0.8	16	1.75	0.16	2	100	27.3	20	5	0.5	12	19.5	1.77	2	4.44	0.4	4.04	15.6	1.41	43.3	Execellent
feb	8.2	96.5	8.77	400	40	3.6	85	14	1.3	20	2.17	0.2	2.1	105	28.6	28	7	0.6	6.1	10.1	0.92	4	8.89	0.81	4.85	18.6	1.69	46.6	Execellent
mar	7.12	83.8	7.61	410	41	3.7	86	14	1.3	16	1.74	0.16	1.9	95	25.9	29	7.3	0.7	11	18.6	1.69	1.87	4.16	0.38	3.57	13.7	1.25	42.7	Execellent
apr	7.9	92.9	8.45	400	40	3.6	75	13	1.1	22	2.39	0.22	3	150	40.9	30	7.5	0.7	6.1	10.1	0.92	2	4.44	0.4	5.18	19.9	1.81	58.2	Good
may	8.23	96.8	8.8	490	49	4.5	108	18	1.6	20	2.17	0.2	3.1	155	42.3	28	7	0.6	12	19.4	1.77	1.86	4.13	0.38	4.49	17.3	1.57	61.7	Good

5.5.2 Fuzzy logic model for stream water quality assessment

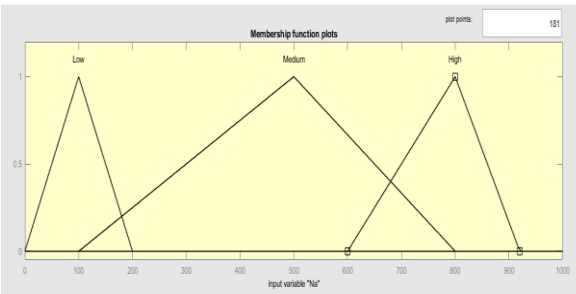
The fuzzy water quality model is developed for determining water quality in the streams or rivers. Various forms of membership functions may be utilized based on the type and needs of the application in consideration. The accurate forecast of the developed fuzzy model usually depends upon the exact number of fuzzy sets used for the mapping the input and output parameters, because it assists in giving further continuity to the "universe of discourse". In the present study, each input quality parameter is separated into three fuzzy subsets. The triangular and trapezoidal membership functions were consigned to the input and output variables as shown in Figure 5.11 and Figure 5.14. The ranges for fuzzy sets have been chosen to determine the water quality of the stream by means of a cumulative index known as "fuzzy water quality index" (FWQI). The parameters are united into linguistic rules using the perception of the binary operator 'AND'. The "minimum" fuzzy operator was used in the study since most of the parameters are self-determining and independent in character. The fuzzy rules were not assigned with any weightage since the whole rule holds equal weightage in order to determine the final WQI.



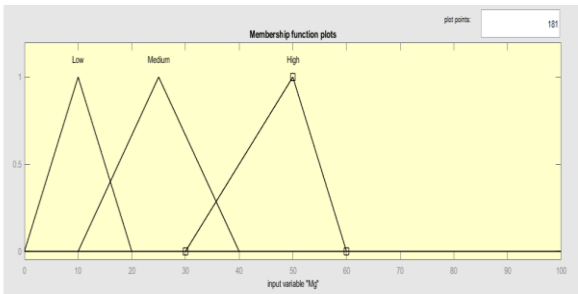
(a) pH



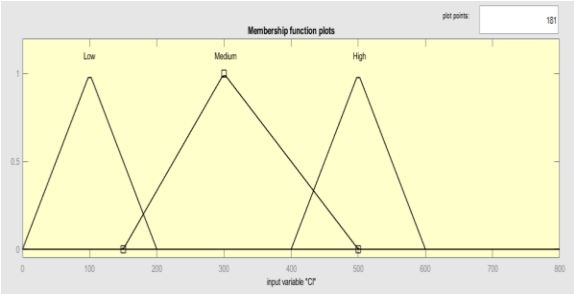
(b) Ec



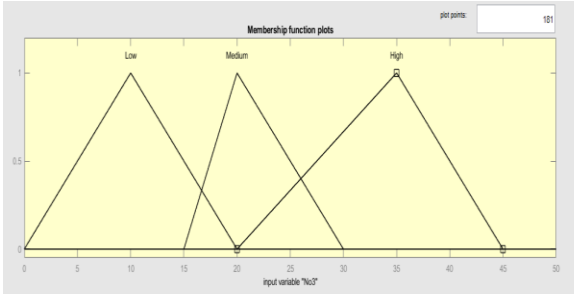
(c) Na



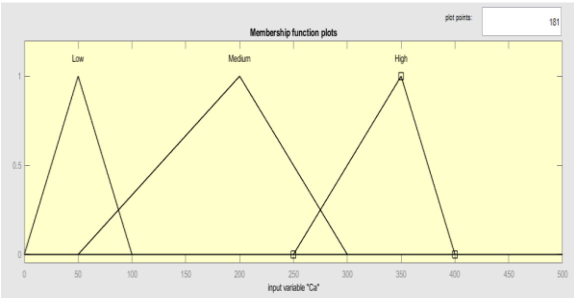
(d) Mg



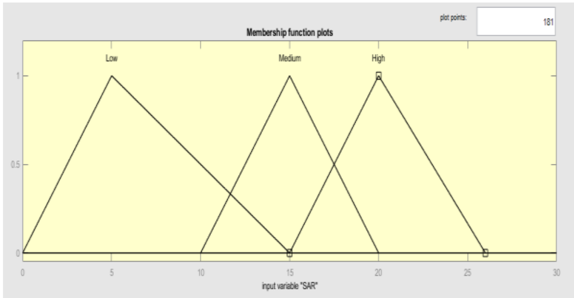
(e) Cl



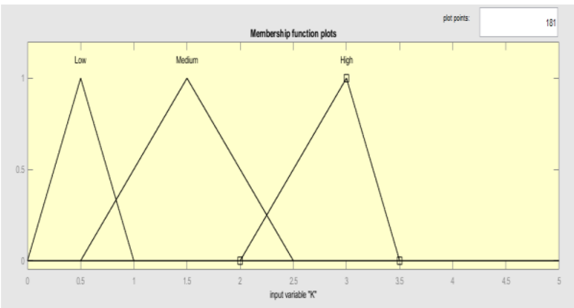
(f) No3



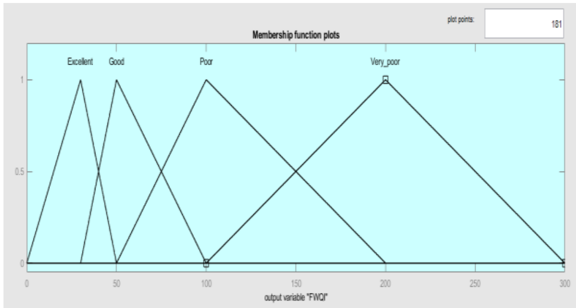
(g) Ca



(h) SAR



(i) K



(j) FWQI

Figure 5.11 Triangular membership functions for input and output parameters

The if-then rules in the rule editor are formulated on the basis of 12 realistic conditions available and hence 12 rules are generated. However, there is no standard method to fix the number of rules. The user has to analyse and formulate the rules according to the application for which the model is developed. Figure 5.12 and Figure 5.13 show the rule editor and rule viewer windows respectively for the developed fuzzy model.

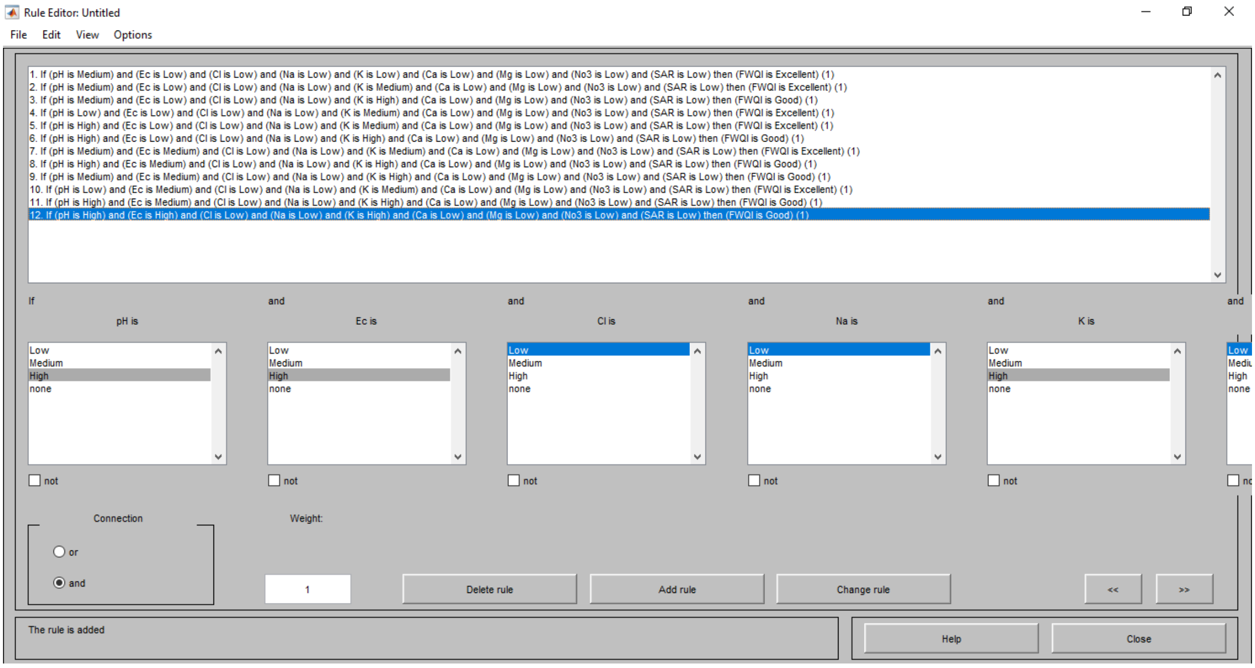
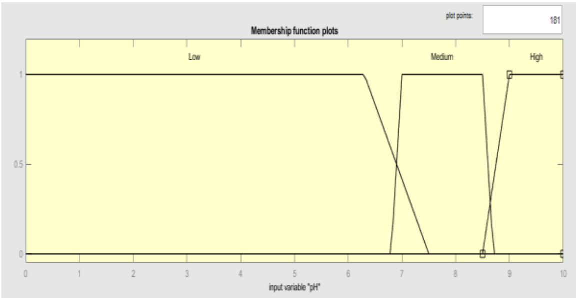
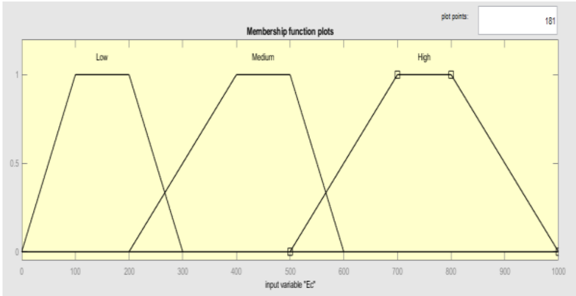


Figure 5.12 Rule editor for triangular membership function

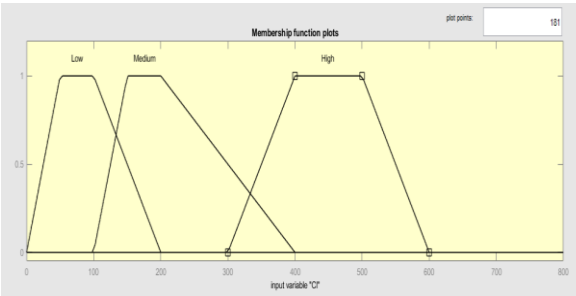
Figure 5.13 Rule viewer for triangular membership function



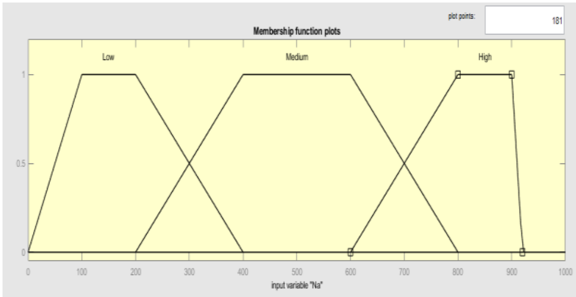
(a) P^H



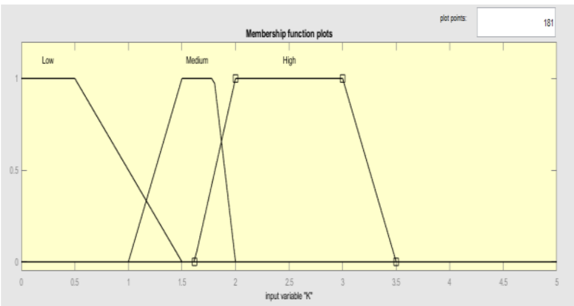
(b) Ec



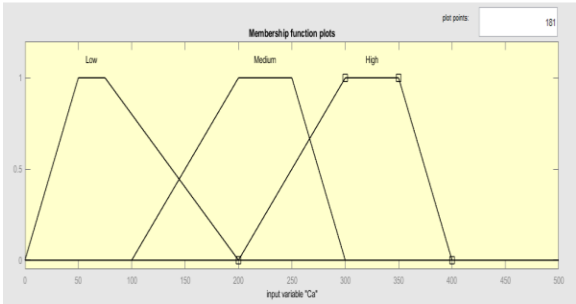
(c) Cl



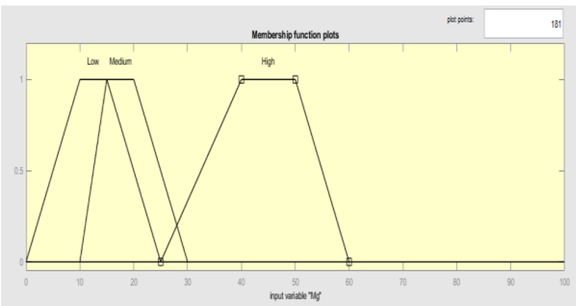
(d) Mg



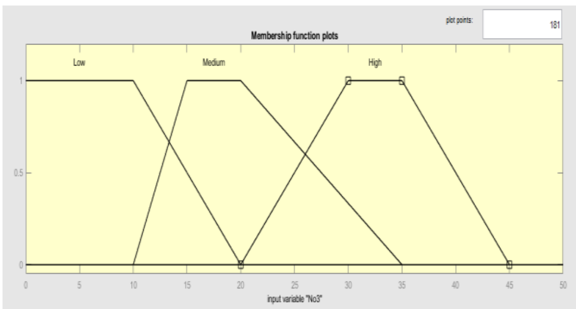
(e) K



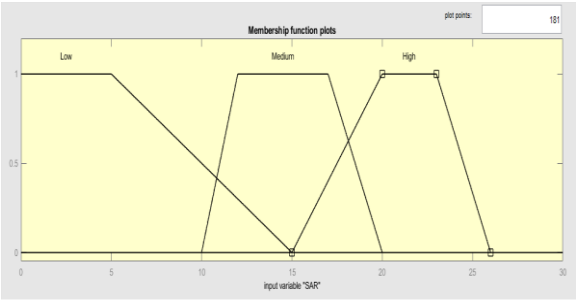
(f) Ca

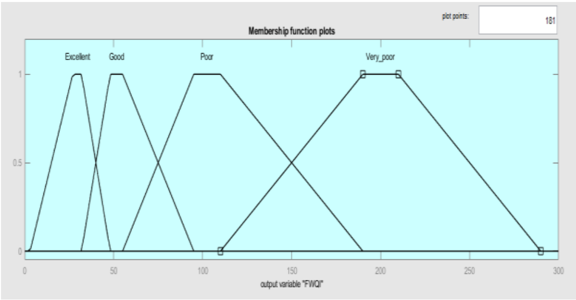


(g) No3



(h) Na





(i) SAR (j) FWQI

Figure 5.14 Trapezoidal membership functions of input and output parameters

The rule editor and rule viewer windows for trapezoidal membership function are as shown in Figure 5.15 and Figure 5.16.

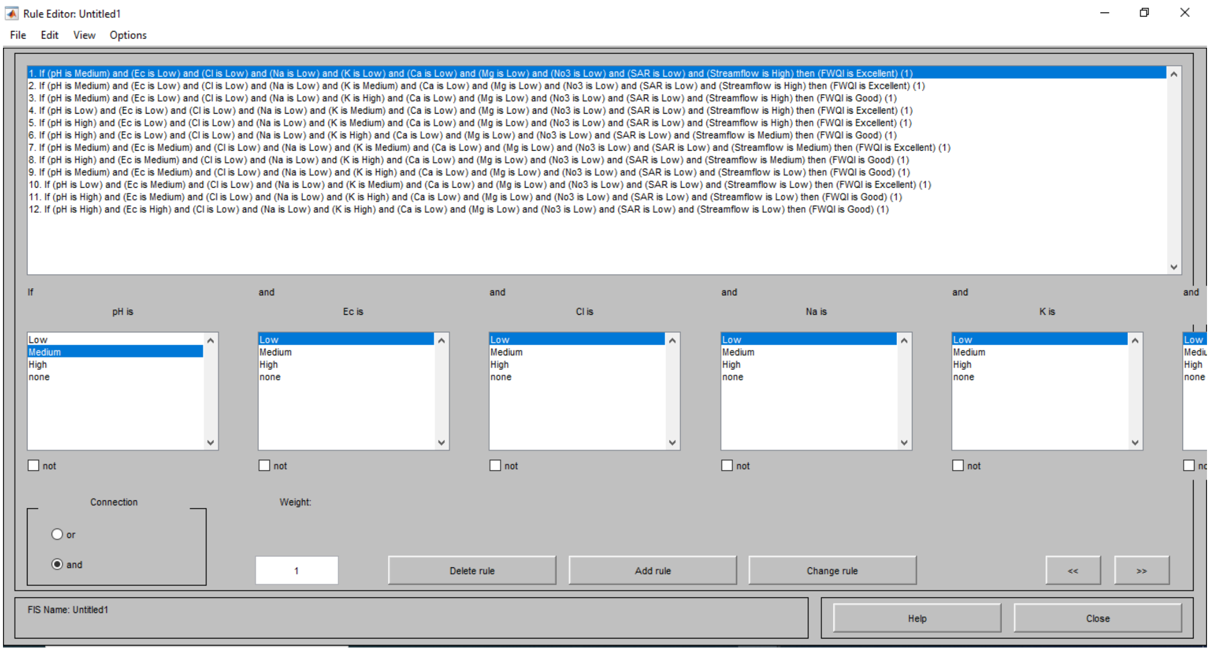


Figure 5.15 Rule editor for trapezoidal membership function

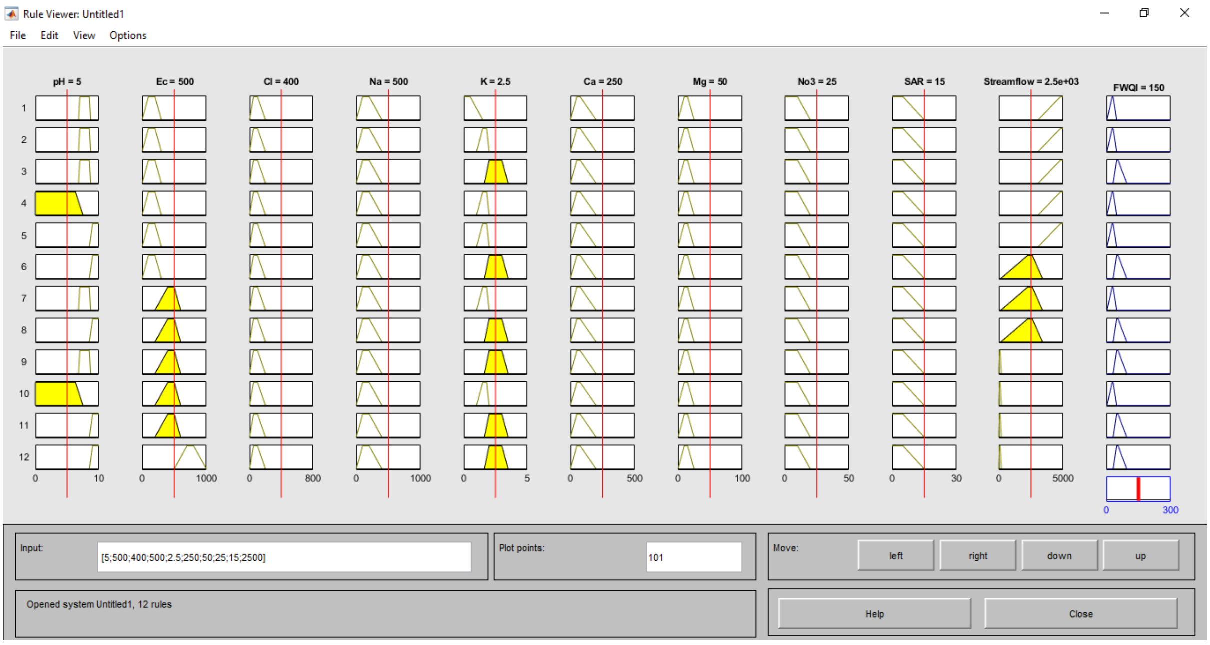


Figure 5.16 Rule viewer for trapezoidal membership function

Further in this study, three defuzzification methods are experimented for evaluation of WQI. They are Centroid method, Bisector method and Mean of Maxima (MOM) method. Evaluation of WQI is carried out both with and without considering stream as one of the inputs to the model so as to get a clear picture of whether streamflow influences the final WQI. The performance of each model is presented in Table 5.15.

Table 5.15 Performance of fuzzy logic model to evaluate WQI

Model	Membership function	Defuzzification method	Performance without streamflow		Performance with streamflow	
			R^2	RMSE (m^3/s)	R^2	RMSE(m^3/s)
M1	Triangular	Centroid	0.6784	20.12	0.6528	21.034
M2		Bisector	**0.7241**	16.28	**0.7042**	15.823
M3		MOM	0.6543	21.23	0.6285	22.894
M4	Trapezoidal	Centroid	0.6482	21.89	0.6089	22.13
M5		Bisector	**0.7326**	16.88	**0.6924**	15.54
M6		MOM	0.614	22.41	0.5826	24.32

From Table 5.15, it is evident that Bisector method of defuzzification outperforms the other two methods for both the membership functions and for both the cases of with and without streamflow. It can be seen that there is no significant change in the model performance when streamflow is considered as one of the inputs. It is difficult to draw a conclusion from the results about the influence of streamflow values on WQI values. Figure 5.17 shows the variation of FWQI values with streamflow values of the same period.

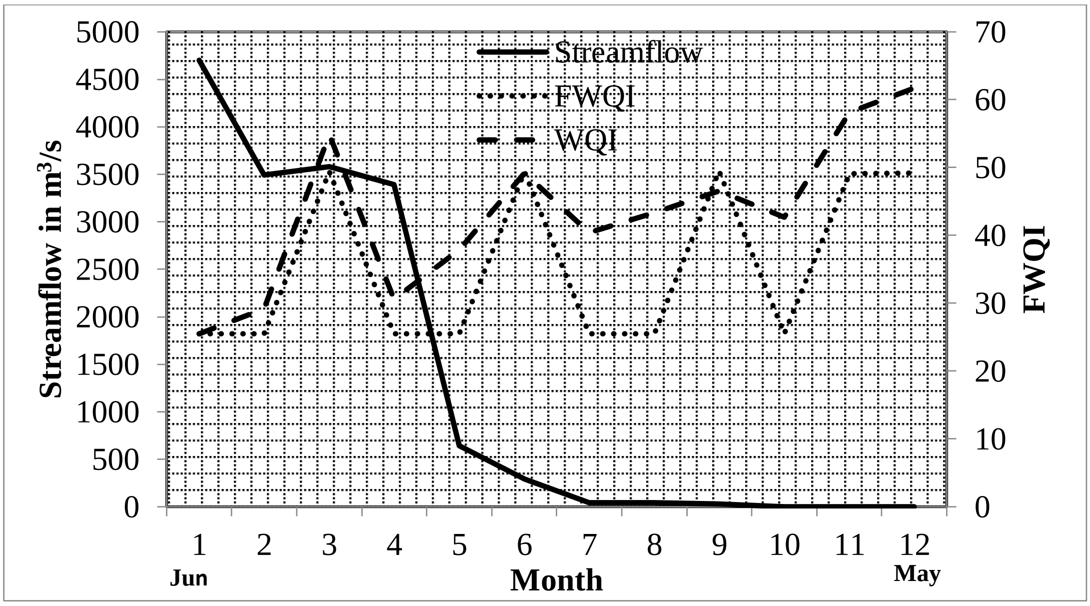


Figure 5.17 Variation of FWQI and streamflow values throughout the year

It can be seen from Figure 5.17 that there no much variation in FWQI along the year. This may be due to the location of sampling point. Khanapur gauging station is at the upstream location of Malaprabha sub-basin, due to which the water quality at that point is in desirable ranges throughout the year. The region is comparatively covered with thick forest at this upstream location and hence less polluted.

5.5.3 Hypothetical simulation of FWQI model

Hypothetical modelling and simulation involve developing a model for various possible conditions for a particular study in cases where we have limited realistic data and resources. A hypothesis is necessarily a system of reasoning that involves those conditions whose reality is yet to be published (Gladun 1997). A few researches can be found in literature regarding hypothetical modelling in water quality management. Mujumdar and Sasikumar (2002) developed a fuzzy risk approach method for seasonal water quality management where they illustrated a fuzzy optimization model with a hypothetical river system. Rehana and Mujumdar (2011) studied the stream water quality response under various hypothetical climate change scenarios.

In our study, we have considered a case of moderately polluted stream as a hypothetical scenario for simulating monthly stream water quality. We have assumed a moderately polluted stream where the input water quality parameters are above their permissible limits by 25%. That means each of the quality parameter falls out of its permissible limit by 25%.Table 5.16lists the weights of each parameter.

Table 5.16 Weights of each parameter for the hypothetical case of a moderately polluted stream

Parameter	Weight
pH	1
Ec	4
Na	4
Mg	1
Cl	4
NO_3	1
Ca	1
SAR	4
K	4

For this case, we have considered the sensitivity of the parameters for irrigational use of water as per IS 2296:1992 where we can find the irrigational water quality standards. Hence the parameters such as Electrical conductivity, sodium, chlorides, sodium adsorption ratio and potassium are considered to fall above the standard limits.

The results of hypothetical simulation of fuzzy water quality model are presented in Table 5.17.

Table 5.17 Results of the hypothetical simulation of stream water quality

Months	Hypothetical Case	
	WQ	FWQ
Jun	Poor	Poor
Jul	Poor	Poor
Aug	Poor	Poor
Sep	Poor	Poor
Oct	Poor	Poor
Nov	Poor	Poor
Dec	Poor	Poor
Jan	Poor	Poor
Feb	Poor	Poor
Mar	Poor	Poor
Apr	Poor	Poor
May	Poor	Poor

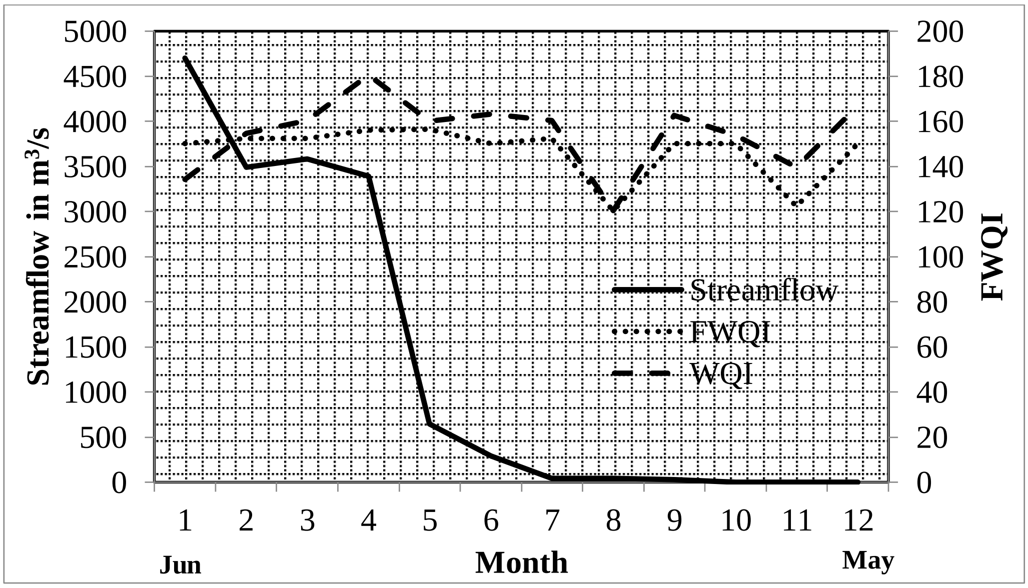


Figure 5.18 Monthly variations of WQI and FWQI with streamflow for Case 2 scenario

From Table 5.17 and Figure 5.18, it is evident that the FWQI model performs well for any given set of input data. This model can provide the WQI directly in terms of class, which can be used as an important tool to determine instant water quality of any stream. It can reduce the time consumption and the computational complexities associated with the calculation of WQI.

5.6 Summary

Autocorrelation analysis is an important step in the Exploratory Data Analysis of time series forecasting. The autocorrelation analysis helps detect patterns and check for randomness while PACF helps to establish number of lags to be considered for modelling. This is not sufficient in case of non-linear models. In order to address the non-linear nature of inputs, outputs and their relationship, hybridization of models is necessary. In the present study, we have used wavelet model to address the noise and non-linearity in the data.

The potential of wavelet method to process the raw streamflow data and denoise the series is evaluated for the application of streamflow forecasting. The results show that wavelet method can improve the performance of forecasting model significantly. Wavelet can be an efficient data pre-processing method for hydrologic forecasting applications. The stream water quality is determined using a fuzzy water quality model which incorporated streamflow as one of the inputs. The results show that streamflow can be linked to the water quality of the stream when both are considered for same time periods. The FWQI model can be simulated for any given set of input data.

Though the forecasting models developed in this study show satisfactory results, the uncertainties are not addressed. Hydrological model uncertainties stem from parameters, model structure, calibration (observation) and input data. In addition to these sources, uncertainties can stem from model initial and boundary conditions (Moges et al. 2021). The cause for underestimation of peaks by all the three forecasting models can be attributed to the uncertainties involved in modelling.

CHAPTER 6

CONCLUSIONS

6.1 General

This chapter presents the conclusions drawn based on the research insights gained from the empirical modelling of streamflow time series and the fuzzy stream water quality modelling. Some of the limitations of the study are listed along with the future scope and recommendations.

6.2 Conclusions

- The performance of SVM, ANFIS, and Fuzzy models to forecast daily streamflow is tested for 1-day, 3-days and 5-days ahead forecasts. The results indicate that the performance of the models significantly decreases with an increase in lead times. The models show high R^2values for 1-day ahead streamflow forecasts, whereas it is low for 3 and 5-days lead time. This is caused by the uncertainty involved in modeling. Also, the historical daily streamflow data used in the study is highly variable on a temporal scale.

- The performance of all the three models significantly increased when the wavelet is coupled with. The WT-fuzzy model outperformed WT-ANFIS and WT-SVM models. However, WT-ANFIS performed better than WT-SVM. The uncertainty involved in the model structure may be one of the reasons for the poor performance of the models.

- In this study, an attempt has been made to relate water quantity with quality by considering streamflow as one of the inputs to the water quality model. A simple fuzzy logic model is developed to evaluate the water quality index.

- WQI is grouped into five classes, namely, "Excellent", "Good", "Poor", "Very Poor", and "Unsuitable". It can be seen from the results that that water quality degrades from monsoon season to dry season. All the parameters except K are within the limits for most of the months. The unacceptable values of K make the water quality poor and very poor for most of the months.

- Two types of membership functions known as 'Triangular' and 'Trapezoidal' are used in the study and the results are consistent for both. Also, three different defuzzification methods are used in which the 'Bisector' method gave comparatively good results for both the membership functions and both the cases with and without streamflow as one of the inputs. The consistency in the results may be due to the location of the sampling point which is upstream of the basin and hence no much change can be seen in the water quality throughout the year.

- The hypothetical simulation was carried out by considering a critical case scenario for determining the irrigation water quality. The results show that the model is capable of handling various sets of input parameters effectively. In order to obtain a complete picture of water quality conditions in the basin, we need to have a minimum of 10 years of river water quality data. The fuzzy model developed in the study can be used to obtain the WQI for any set of input variables in the future, thereby reducing the computational complexities of WQI.
- The results indicate that streamflow can be considered as one of the inputs to determine the WQI. But more detailed studies have to be carried out related to dilution and parameter concentrations for various values of streamflow throughout the year for a better analysis of stream water quality and quantity. Also, detailed flood hydrograph and return periods for various floods for the basin must be studied in order to know the exact ranges of input streamflow.

6.3 Contributions from the study

- The importance of data pre-processing is illustrated in the study so as to make hydrologic modelling and forecasting more efficient. The noise in the raw data can be reduced using data pre-processing techniques like wavelet transform.
- Choice of mother wavelet is one of the crucial steps in wavelet analysis. This study demonstrated that the mother wavelet function is sensitive to the length of the data and gauging station location.
- Wavelet method is an efficient tool for data pre-processing. The results of the study show that there is a significant increase in forecasting accuracy when the model was coupled with wavelet.
- The stream water quality and quantity can be integrated by means of a FWQI and the stream water quality can be determined for any given set of data using the fuzzy model. This can save the computational time and aids the efficient operation and management of water resources.

6.4 Limitations of the study

- The three soft computing models developed to forecast the daily streamflow time series show underestimation of peak flows.
- The results of forecasting show some negative streamflow values which indicate the uncertainties in modelling.
- The water quality research in this case study suffered lack of data and experimentations.
- The uncertainties in modelling are not addressed properly in this research.

6.5 Future scope of research

- The study compared only three soft computing methods, namely, SVM, ANFIS and Fuzzy Logic. Other methods like genetic algorithms, particle swarm optimization techniques, etc., can be compared so as to provide the best method for forecasting.
- The non-availability of water quality data for a longer period is a major limitation of the study. Fuzzy model has to be trained and tested for at least 10 years of water quality data.
- The Khanapur station is located at the upstream of the basin where the water quality is relatively good. This particular data failed to convey the effectiveness of the fuzzy model to assess the stream water quality. Hence, we have developed a hypothetical scenario to check the model with variations in the data. The model has to be checked for any extremities of the data with a longer time span.
- Stream water quality and quantity related aspects like dispersion ratio, a detailed flood hydrograph analysis, terrain information, etc., have to be analysed and applied in to the model.

REFERENCES

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

Adamowski, J., and Sun, K. (2010). "Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds."*J Hydrol* 390(1):85–91. https://doi.org/10.1016/j.jhydrol.2010.06.033.

Adamowski, J., and Sun, K. (2010). "Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds."*J Hydrol.,* 390(1):85–91. https://doi.org/10.1016/j.jhydrol.2010.06.033.

Adeloye, A.J., and Montaseri, M. (2002). "Preliminary streamflow data analyses priorto water resources planning study."*Hydrological Sciences Journal,* **47(5):** 679-692.

Ahmed, Y., Huang, Y., Martin, A., and Otten B. (2015). "Assessment of the Relation between Water Quality and Water Quantity for International Metropolitan Cities". *World Environmental and Water Resources Congress,* 2015: 669, Floods, Droughts, and Ecosystems © ASCE 2015.

Asefa, T., Kemblowski, M. W., McKee, M., and Khalil, A. (2006). "Multi-time scale stream flow prediction: The support vector machines approach." *Journal of Hydrology,* 318: 7–16.

Aussem, A., and Murtagh, F. (1997)."Combining Neural Network Forecasts on Wavelet-transformed Time Series" *Connection Science,* 9:1, 113-122, DOI: 10.1080/095400997116766

Bae, D. H., Jeong, D. M., and Kim, G. (2007)."Monthly dam infow forecasts using weather forecasting information and neuro-fuzzy technique."*Hydrol Sci J,* 52:99–113.

Bardossy, A. (1996). "The use of fuzzy rules for the description of elements of the hydrological cycle."*Ecol. Modell.* 85 (1): 59–65, doi:10.1016/0304-3800 (95)00011-9

Behzad, M. K., Asghari, M., Eazi, M., and Palhang. (2009). "Generalization performance of support vector machines and neural networks runoff modeling." *Expert System with Applications,* 36: 7624-7629.

Bhadra, A., Bandyopadhyay, A., Singh, R., and Raghuwanshi, N. S. (2010). "Rainfall-Runoff Modeling: Comparison of Two Approaches with Different Data Requirements."*Water Resources Management,* 24(1), 37–62. doi:10.1007/s11269-009-9436-z

Bhat, S. A., Meraj, G., and Pandit, A. K. (2015)."Assessing the Influence of Stream-flow and Precipitation Regime on Water Quality of the Major Inflow Stream of Wular Lake in Kashmir Himalaya.." doi:10.1007/s12517-015-2083-1.

BIS (Bureau of Indian Standards).(2002). "Tolerance Limits of Selected Water Quality Parameters for Inland Surface Water Prescribed for Different uses by Bureau of Indian Standards.", Bureau of Indian Standards, New Delhi.

BIS (Bureau of Indian Standards).(2004). "Drinking Water Specification (First Revision)," IS: 10500: 2004-05. Bureau of Indian Standards, New Delhi.

Bodo, B., Unny, T. E. (1983)."Sampling strategies for mass discharge estimation."*J Environ Eng*., 109:812–828.

Bourdin, D. R., and Fleming S. W., and Stull, R. B. (2012). "Streamflow Modelling: A Primer on Applications, Approaches and Challenges." Atmosphere-Ocean, 50:4, 507-536, DOI: 10.1080/07055900.2012.734276

Budu, K. (2013). "Comparison of Wavelet-Based ANN and Regression Models for Reservoir Inflow Forecasting." Journal of Hydrologic Engineering, Vol. 19, Issue 7 © 2013 American Society of Civil Engineers. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000892

Budu, K., Rao, S., Ramji, Y., and Chandra, N. P. (2012). "Wavelet neural network model for river flow time series." *Proceedings of the Institution of Civil Engineers - Water Management, 165(8), 425–439.* doi:10.1680/wama.10.00092

Chandwani, V., Vyas, S. K.,Agrawal, V., and Sharma, G.(2015)."Soft computing approach for rainfall-runoff modelling : A review."International Conference on Water Resources, Coastal and Ocean,*Aquat. Procedia,* **4,** 1054–1061.

Chang, F. J., Chiang, Y. M., Tsai, M. J., Shieh, M. C., Hsu, K. L., and Soroosh, S. (2014)."Watershed rainfall forecasting using neuro-fuzzy networks with the assimilation of multi-sensor information."*J Hydrol.*, 508:374–384.

Chang, F., and Chen, Y. (2001). “A counterpropagation fuzzy-neural network modeling approach to real time streamflow prediction.”*J. Hydrol.* 245 (1–4): 153–164, doi:10.1016/S0022-1694(01)00350-X

Chaturvedi, D. (2008). “Soft Computing Techniques and its Applications in Electrical Engineering.” *In: Studies in Computational Intelligence.* Springer, Berlin, 1-10.

Chau, K.W., Wu, C. L., and Li, Y. S. (2005). “Comparison of several flood forecasting models in Yangtze River.”*J. Hydrol. Eng.* 10 (6): 485–491, doi:10.1061/ (ASCE)1084-0699(2005)10:6(485)

Chen, J., and Adams, B. J. (2006).“Integration of artificial neural networks with conceptual models in rainfall-runoff modeling.”*Journal of Hydrology*, 318(1-4), 232-249, Copyright 2006, with permission from Elsevier.

Chong, K. L., Lai, S. H., and El-Shafie, A. (2019). “Wavelet Transform Based Method for River Stream Flow Time Series Frequency Analysis and Assessment in Tropical Environment.” Water Resour Manage 33, 2015–2032. https://doi.org/10.1007/s11269-019-02226-7

Clement, A. (2001).“Improving uncertain nutrient load estimates for Lake Balaton.”*Water Sci Technol,* 43:279–286.

Coulibaly, P., and Burn, D. H. (2004).“Wavelet analysis of variability in annual Canadian streamflows.”*Water Resour. Res.,* 40, W03105, doi:10.1029/2003WR002667.

Daubechies, I. (1992). “Ten lectures on wavelets”: SIAM.

Dawson, C.W., and Wilby, R. L. (2001). “Hydrological modelling using artificial neural networks.”*Prog. Phys. Geog*. 25 (1): 80–108, doi:10.1177/ 030913330102500104

Deka, P. C., and Chandramouli, V. (2009).“Fuzzy neural network modeling of reservoir operation.”*J Water Resour. Plan Manag.,* 135:5–12.

Deka, P., and Chandramouli, V. (2003).“A fuzzy neural network model for deriving the river stage-discharge relationship.”*Hydrol. Sci. J.*48:197–209.

Dongjie, W., Wang, J., and Teng, Y. (2004).“Prediction of undergroundwater levels using wavelet decompositions and transforms.” *Journal of Hydro-Engineering* 5(4): 34–39

Dunca, A. M. (2018). “Water Pollution and Water Quality Assessment of Major Transboundary Rivers from Banat (Romania).”*HindawiJournal of Chemistry*,Volume 2018, Article ID 9073763, 8 pages https://doi.org/10.1155/2018/9073763.

Ejigu, M. T. (2021). “Overview of water quality modelling”, Cogent Engineering, 8:1, 1891711, DOI: 10.1080/23311916.2021.1891711

Elshorbagy, A., Simonovic, S. P., Panu, U. S. (2002). “Noise reduction in chaotic hydrologic time series: Facts and doubts.” Journal of Hydrology, 256(3-4):147-165. DOI:10.1016/S0022-1694(01)00534-0

FAO, (1985). “Water Quality for Agriculture. Irrigation and Drainage.” Paper No. 29 Rev.1, Rome, 182 pp.

Figueiredo, K., Barbosa, C. R. H., Cruz, A. D., Vellasco, M., Pacheco, M., Conteras, R. (2007).“Neural networks for inflow forecasting using precipitation information.”*Lect. Notes Comput. Sci*. 4570 :552-561.

Fletcher, R.(1987). in: Practical Methods of Optimization, second ed., Wiley, New York.

Flood, I., and Kartam, N. (1994). “Neural networks in civil engineering. I: Principles and understanding.”*Journal of Computing in Civil Engineering*, ASCE, 8(2), 131-148.

Flood, I., and Kartam, N. (1994). “Neural networks in civil engineering. II: Systems and application.”*Journal of Computing in Civil Engineering*, ASCE, 8(2), 149-162.

Ganoulis, J. (2009). “Risk Analysis of Water Pollution”. Second, Revised and Expanded Edition. Copyright 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, ISBN: 978-3-527-32173-5.

Gladun, V. P. (1997). “Hypothetical Modeling: Methodology and Application”. *Cybernetics and Systems Analysis*, Vol. 33, No. 1.

Graps, A. (1995). "An introduction to wavelets," in IEEE Computational Science and Engineering, vol. 2, no. 2, pp. 50-61, Summer, doi: 10.1109/99.388960.

Guo, Y., Markus, M., and Demmissie, M. (2002).“Uncertainty of nitrate-N load computations for agricultural watersheds.” *Water Resour Res*, 38(10):3–1, 3-12.

Haan, C.T. (2002). “Statistical Methods in Hydrology.” 2nd edition. Iowa State University Press, Iowa, USA.

Hadi, S. J., and Tombul, M. (2018). “Forecasting Daily Streamflow for Basins with Different Physical Characteristics through Data-Driven Methods.” *Water Resour. Manag.* 32(10): 3405–3422.

Han D., Chan, L., and Zhu, N. (2007). “Flood forecasting using support vector machines.” *Journal of hydroinformatics*. 9(4), 267-276.

Han, J. C., Huang, G. H., Zhang, H., Zhuge, Y. S., and He, L. (2012).“Fuzzy constrained optimization of eco-friendly reservoir operation using self-adaptive genetic algorithm: A case study of a cascade reservoir system in the Yalong River, China.”*Ecohydrology*5:768–778.

Haykin, S. (2009).“Neural networks and learning machines, *3rd edn. Pearson,* Upper Saddle River

He, Z., Wen, X., Liu, H., and Du, J. (2014).“A comparative study of the artifcial neural network, adaptive neuro-fuzzy inference system and support vector machine for forecasting river fow in the semiarid mountain region.”*J Hydrol* 509:379–386.

Hipel, K. W. (1985), “Stochastic research in multivariate analysis”. Keynote paper presented at the Fourth International Hydrology Symposium on “Multivariate analysis of Hydrologic Processes”, held at Colorado State University, Fort Collins, Colorado, July 15–17, 1985.

Honorato, A. G. S., Silva, G. B. L., and Santos, C. A. G. (2018). “Monthly streamflow forecasting using neuro-wavelet techniques and input analysis.” Hydrological Sciences Journal, 63:15-16, 2060-2075, DOI: 10.1080/02626667.2018.1552788

Horton, R. K.(1965). “An index number system for rating water quality.”*Journal of the Water Pollution Control Federation,*37 (3), 300–306.

Hsieh, C. J., Chang, K. W., Lin, C. J., Keerthi, S. S., and Sundararajan, S. (2009). “A dual coordinate descent method for large-scale linear SVM.” *In Proceedings of the 25th International Conference on Machine Learning (ICML’08)*, 408–415.

Jain, A., and Kumar, A. M. (2007). “Hybrid neural network models for hydrologic time series forecasting.”, 7(2), 585–592. doi:10.1016/j.asoc.2006.03.002

Jang, J. S. R. (1993). “ANFIS: adaptive network based fuzzy inference systems.”*IEEE Trans. Syst.,* Man Cybernet. 23 (3):665–685.

Jang, J. S. R. (1997). “Self learning fuzzy controllers based on temporal back propagation.”*IEEE Trans. Neural Networks*. 3 (5):714–723.

Jeong, C., Shin, J. Y.,Kim, T., and Heo, J. H. (2012).“Monthly precipitation forecasting with a neuro-fuzzy model.”*Water Resour. Manag.,* 26:4467–4483.

Kangabam, R. D., Bhoominathan, S. D., Kanagaraj, S., and Govindaraju, M. (2017). “Development of a water quality index (WQI) for the Loktak Lake in India.”*Appl Water Sci*, 7:2907–2918 DOI 10.1007/s13201-017-0579-4.

Keskin, M. E., Terzi, O., Taylan, D. (2004).“Fuzzy logic model approaches to daily pan evaporation estimation in western Turkey.”*Hydrologic Sci. J.,* 49(6), 1001–1010.

Keskin, M. E., Taylan, D., Terzi, Ö. (2006). “Adaptive neural-based fuzzy inference system (ANFIS) approach for modelling hydrological time series.” *Hydrological Sciences Journal, 51(4), 588–598.* doi:10.1623/hysj.51.4.588

Kisi, O., and Cimen, M. (2011). “A wavelet-support vector machine conjunction model for monthly streamflow forecasting.”*J. Hydrol*. 399(1–2), 132–140.

Kişi, Ö., Nia, A. M., Gosheh, M. G., Tajabadi, M. R. J., and Ahmadi, A. (2012). “Intermittent streamflow forecasting by using several data driven techniques.”*Water Resour Manag.* 26:457–474.

Koza, J.R. (1992). “Genetic programming: On the programming of computers by means of natural selection.” Cambridge, Massachusetts: MIT Press.

Kükrer, S., and Mutlu, E. (2019). “Assessment of surface water quality using water quality index and multivariate statistical analyses in Saraydüzü Dam Lake, Turkey.”*Environ Monit Assess*, 191:71 https://doi.org/10.1007/s10661-019-7197-6

Lan, Y. H. (2001). “A genetic-algorithm based automatic model calibrator for the UBC watershed model.” M.S. thesis, Dep. of Civil Engineering, Univ. of British Columbia, Vancouver, British Columbia.

Leelavathy, K. R., Nageshwaran, V., and Nirmala, V.(2017).“ANN Approach to Analysis the Quality of River Water -A Case Study of Chunnambar River Pondicherry.”*International Journal of Engineering Research & Technology (Ijert) Iconnect* – 2017 (Volume 5 – Issue 13).

Liao, S. H. (2005). “Expert system methodologies and applications—a decade review from 1995 to 2004.”*Expert Syst. Appl*. 28 (1): 93–103, doi:10.1016/j.eswa.2004.08.003

Liong, S., and Sivapragasam, C. (2002). “Flood stage forecasting with support vector machines.” *J. of. Water Resour.* Assoc. 38 (1): 173–186, doi:10.1111/j.1752-1688.2002.tb01544.x

Lohani, A. K., Kumar, R., and Singh, R. D. (2012) “Hydrological time series modeling : A comparison between adaptive neuro-fuzzy , neural network and autoregressive techniques.”*J. Hydrol.***442–443,** 23–35.

Luger, G. F. (2005). “Artificial intelligence: Structures and strategies for complex problem solving.” (5th ed.). Boston: Addison-Wesley.

Machiwal, D., and Jha, M. K. (2012). “Hydrologic time series analysis: theory and practice.” Springer Science & Business Media.

Machiwal, D., and Jha, M.K. (2006). “Time series analysis of hydrologic data for waterresources planning and management: A review.”*Journal of Hydrology andHydromechanics,* **54(3):** 237-257.

Machiwel, D., and Jha, M. K. (2012). “Current Status of Time Series Analysis.”*In Hydrological Sciences,* © Capital Publishing Company doi:10.1007/978-94-007-1861-6

Machiwel, D., and Jha, M. K. (2012).“Methods for Time Series Analysis.”*In Hydrological Sciences,* © Capital Publishing Company doi:10.1007/978-94-007-1861-6

Magar, R. B., and Jothiprakash, V. (2011). “Intermittent reservoir daily-inflow prediction using lumped and distributed data multi-linear regression models.”*J. Earth Syst. Sci., 120(6), 1067–1084.* doi:10.1007/s12040-011-0127-9.

Mahabir, C., Hicks, F. E., and Faytek, A. R. (2003). “Application of fuzzy logic to forecast seasonal runoff.”*Hydrol. Process*. 17 (18): 3749–3762, doi:10.1002/hyp.1359

Mallat, S. (1998) “A Wavelet Tour of Signal Processing.” Academic Press, Cambridge.

Mamdani, E. H., and Assilian, S. (1975). “An experiment in linguistic synthesis with a fuzzy logic controller.”*Int. J. Man Mach. Stud.* 7 (1):1–13.

Mehr, A. D., Ercan, K., and Ehsan, O. (2013). "Streamflow prediction using linear genetic programming in comparison with a neuro-wavelet technique." *Journal of Hydrology, 505(), 240–249.* doi:10.1016/j.jhydrol.2013.10.003

Moges, E., Demissie, Y., Larsen, L., and Yassin, F. (2021). "Review: Sources of Hydrological Model Uncertainties and Advances in Their Analysis." Water, MDPI, 13(1), 28; https://doi.org/10.3390/w13010028

Mudbhatkal, A., Raikar, R. V., Venkatesh, B., and Mahesha, A. (2017). "Impacts of Climate Change on Varied River-Flow Regimes of Southern India."*J. Hydrol. Eng.,* 22(9):05017017.

Mujumdar, P. P., and Sasikumar, K. (2002)."A fuzzy risk approach for seasonal water quality management of a river system."*Water Resour. Res.,*38:5–9.

Nash, J. E., and Sutcliffe, J. V. (1970). "River flow forecasting through conceptual models part I—A discussion of principles."*J Hydrol.,* 10(3):282–290.

Nayak, P. C., Sudheer, K. P., Rangan, D. M., and Ramasastri, K. S. (2004) "A neuro-fuzzy computing technique for modeling hydrological time series."**291,** 52–66.

Nourani, V., Baghanam, A. H., Rahimi, A. Y., and Nejad, F. H. (2014)b."Evaluation of wavelet-based de-noising approach in hydrological models linked to artificial neural networks." In: Islam T, Srivastava PK, Gupta M, Zhu X, Mukherjee S (eds) *Computational intelligence techniques in earth and environmental sciences.* Springer Netherlands, Dordrecht, pp 209–241.

Nourani, V., Hosseini, A., Adamowski, J., and Kisi, O. (2014)."Applications of hybrid wavelet– Artificial Intelligence models in hydrology : A review."*J. Hydrol.***514,** 358–377.

Okkan, U. (2012)."Using wavelet transform to improve generalization capability of feed forward neural networks in monthly runoff prediction."*Sci. Res. Essays.* 7(17) (2012) 1690-1703.

Oladipo, J. O., Akinwumiju, A. S., Aboyeji, O. S., and Adelodun, A. A. (2021). "Comparison between fuzzy logic and water quality index methods: A case of water quality assessment in Ikare community, Southwestern Nigeria." *Environmental Challenges, (3), 100038.* doi:10.1016/j.envc.2021.100038.

Özger, M. (2010).“Comparison of fuzzy inference systems for streamflow prediction.”*Hydrological Sciences Journal.* 54, 261–273.

Özger, M., Mishra, A. K., and Singh, V. P. (2012).“Long Lead Time Drought Forecasting Using a Wavelet and Fuzzy Logic Combination Model: A Case Study in Texas.”*J. Hydrometeorol.*13:284–297.

Partal, T., and Kişi, Ö. (2007).“Wavelet and neuro-fuzzy conjunction model for precipitation forecasting.”*J. Hydrol.*342:199–212.

Partal, T., and Kişi, Ö. (2007).“Wavelet and neuro-fuzzy conjunction model for precipitation forecasting.”*J Hydrol.,* 342:199–212

Peng, T., Zhou, J., Zhang, C., Fu, W. (2017). “Streamflow Forecasting Using Empirical Wavelet Transform and Artificial Neural Networks”, Water, MDPI, 9, 406; doi:10.3390/w9060406

Platt, J. C. (1998). “Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines.” Technical Report MSR-TR-98-14.

Pramanik, N., and Panda, R. K. (2009).“Application of neural network and adaptive neuro-fuzzy inference systems for river fow prediction.”*Hydrol. Sci. J.,* 54:247–260.

Raghavendra, N. S., and Deka, P. C. (2014). “Support vector machine applications in the field of hydrology : A review.”*Appl. Soft Comput. J.,* 19, 372–386.

Rajaee, T., Mirbagheri, S.A., Nourani, V. (2010). “Prediction of daily suspended sediment load using wavelet and neurofuzzy combined model.” *Int. J. Environ. Sci. Technol.* **7,** 93–110. https://doi.org/10.1007/BF03326121.

Ramakrishnaiah, C. R., Sadashivaiah, C., Rangana, G. (2009).“Assessment of Water Quality Index for the Groundwater in Tumkur Taluk, Karnataka State, India.”*E-Journal ofChemistry* http://www.e-journals.net, 6(2), 523-530.

Rao, A.R., Hamed, K.H., and Chen, H.-L. (2003). “Nonstationarities in Hydrologic andEnvironmental Time Series.”*Water Science and Technology Library,* **45**, 392 pp.

Rasouli, K., Hsieh, W. W., and Cannon, A. J.(2012). “Daily streamflow forecasting by machine learning methods with weather and climate inputs.”*J. Hydrol.* 414–415: 284–293, doi:10.1016/j.jhydrol.2011.10.039

Raychaudhuri, M., Raychaudhuri, S., Dhal, S., Kumar, A., and Jena, S. K. (2011).“Groundwater quality along Daya river for irrigation use.”In : Workshop on Ground Water Development and Management Prospect in Odisha (March 7th, 2011) (Eds D. P. Pati, P. K. Mahapatra, D. N. Mandal, C. Maohanty and A. Chowdhury). Central Ground Water Authority & Central Ground Water Board, SE Region, Ministry of Water Resources, GOI. 111 – 121.

Raychaudhuri, M., Raychaudhuri, S., Kumar, A., Jena, S. K., and Choudhury, A. (2014).“Water Quality Index to Monitor Water Quality of Rushikuliya Canal Command Area for Irrigation and Potable Use.”**In** Volume of Papers. One Day Workshop on Augmentation and Conservation of Groundwater Resources of Odisha (Eds: P. K. Mohapatra, D. N. Mandal., A. Choudhury, and G. Vijayakumar), Central Ground water Authority and Central Ground water Board, South Eastern Region, Bhubaneswar, pp. 140-147.

Rehana, S., and Mujumdar, P. P. (2011). “River water quality response under hypothetical climate change scenarios in Tunga-Bhadra river, India”. *Hydrol. Process.* DOI: 10.1002/hyp.8057.

Robertson, D. M., Roerisch, E. D. (1999).“Influence of various water quality sampling strategies on load estimates for small streams.”*Water Resour. Res.*, 35:3747–3759.

Russell, S.O., and Campbell, P. F. 1996). “Reservoir operating rules with fuzzy programming.”*J. Water Resour. Plann. Manage*. 122 (3): 165–170.

Ru-zhong, L. I. (2007).“Fuzzy Model for Two-Dimensional River Water Quality Simulation Under Sudden Pollutants Discharged.”*Journal of Hydrodynamics*, **19,** 434–441.

Sahay, R. R., and Sehgal, V. (2014).“Wavelet-ANFIS models for forecasting monsoon flows a Case study of the Gandak River (India).”*Water Resour.*41:574–582.

Sahoo, M. M. (2014). “Analysis And Modelling Of Surface Water Quality In River Basins.”, Dissertation, Master of Technology, Department of Civil Engineering, National Institute Of Technology Rourkela-769008.

Sang, Y., Singh, V. P., Sun, F., Chen, Y., Liu, Y., and Yang, M. (2016). “Wavelet-Based Hydrological Time Series Forecasting.”*J. Hydrol. Eng*. @ASCE. 21(5), 1–5.

Santos, C. A. G., and Silva, G. B. L. (2014). “Daily streamflow forecasting using a wavelet transform and artificial neural network hybrid models.” Hydrological Sciences Journal, 59:2, 312-324, DOI: 10.1080/02626667.2013.800944

Santos, G. C. A., and Silva, G. B. L. (2014). “Daily streamflow forecasting using a wavelet transform and artificial neural network hybrid models.” *Hydrological Sciences Journal, 59(2), 312–324.* doi:10.1080/02626667.2013.800944

Schindler, D. W. (1978).“Factors regulating phytoplankton production and standing crop in the world’s freshwaters.”*Limnol Oceanogr,* 23:478–486.

Seo, Y., and Kim, S. (2016). “River stage forecasting using wavelet packet decomposition and data-driven models.”*Procedia Eng.* 154:1225–1230. doi:10.1016/j.proeng.2016.07.439

Seo, Y., Singh, V., Kisi, O., Kim, S., and Parasuraman, K.(2016). “River stage forecasting using wavelet packet decomposition and machine learning models.”*Water Resour Manage.* 30:4011–4035.

Shiri, J., and Kisi, O.(2010). “Short-term and long-term stream-flow forecasting using a wavelet and neuro-fuzzy conjunction model.”*J. Hydrol.,* 394 (3–4), 486–493.

Shu, C., and Burn, D. H. (2004). “Homogeneous pooling group delineation for flood frequency analysis using a fuzzy expert system with genetic enhancement.”*J. Hydrol*. 291 (1–2): 132–149, doi:10.1016/j.jhydrol.2003.12.011

Shumway, R. H., and Stoffer, D. S. (2010). “Time series analysis and its applications: with R examples.” Springer Science & Business Media.

Singh, K. P., Malik, A., and Sinha, S. (2005). “Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques—a case study.” *Analytica Chimica Acta, 538(1-2), 355–374.* doi:10.1016/j.aca.2005.02.006

Solomatine, D.P., and Ostfeld, A.(2008). “Data-driven modelling: some past experiences and new approaches.”*Journal of Hydroinformatics,* 10(1),3–22..http://dx.doi.org/10.2166/hydro.2008.015.

Subramani, T., and Ismail, S. M. (2015). “Develop To Model Assess Of River Basins In Tamilnadu.”, *International Journal of Application or Innovation in Engineering & Management* (IJAIEM), Volume 4, Issue 5, May 2015, 264-273.

Sunkad, B. N. (2013).“Water quality of Malaprabha River with reference to physicochemical Factors near Khanapur Town of Belgaum district.”*International journal of innovative research in science, engineering and technology,* Vol. 2, issue 4, pp 930-936

Surendra, H. J., and Deka, P. C. (2014) “Development of a Fuzzy Logic Based Model using different Membership and Rules criteria for predicting water consumption using Climatic variables.” *Soft computing,*5:1293–1301.

Takagi, T., and Sugeno, M. (1985). “Fuzzy identification of systems and its application to modeling and control.” IEEE Trans. Syst., Man Cybernet. 15 (1):116–132.

Talei, A., Chua, L. H. C., and Wong, T. S. W. (2010).“Evaluation of rainfall and discharge inputs used by adaptive network-based fuzzy inference systems (ANFIS) in rainfall–runof modeling.”*J Hydrol.,* 391:248–262.

Talei, A., Chua, L. H. C., Quek, C., and Jansson, P. E. (2013).“Runof forecasting using a Takagi–Sugeno neuro-fuzzy model with online learning.”*J. Hydrol.,* 488:17–32.

Tayfur, G., and Brocca, L. (2015). “Fuzzy Logic for Rainfall-Runoff Modelling Considering Soil Moisture.” *Water Resources Management, 29(10), 3519–3533.* doi:10.1007/s11269-015-1012-0

Thornthwaite, C. W. (1948). “An approach toward a rational classification of climate.”*Geogr. Rev.,* 38(1):55–94.

Tiwari, M. K., and Chatterjee, C. (2010). “Development of an accurate and reliable hourly flood forecasting model using wavelet-bootstrap-ANN (WBANN) hybrid approach.”*J. Hydrol*., 394(3–4), 458–470.

Torrence, C., and Compo, G. P. (1998). “A practical guide to wavelet analysis.” J. *of Meteorol. Sci.,* 79(1):61–78.

Tripathi, S., Srinivas, V. V., and Nanjundiah, R. S.(2006). “Downscaling of precipitation for climate change scenarios: A support vector machine approach.”*J. Hydrol.* 330 (3–4): 621–640, doi:10.1016/j. jhydrol.2006.04.030

Tsukamoto, Y. (1979). “An Aapproach to Fuzzy Reasoning Method.” In: Gupta M, Ragade R, Yager R. (Eds.). *Advances in Fuzzy Set Theory*. North-Holland. Amsterdam.

Vanni, M. J., Renwick, W. H., Headworth, J. L., Auch, J. D., and Schaus, M. H. (2001).“Dissolved and particulate nutrient flux from three adjacent agricultural watersheds: a five year study.”*Biogeochemistry,* 54(1):85–114.

Vieux, B. E., and Moreda, F. G. (2003).“Nutrient loading assessment in the Illinois River using a synthetic approach.” *J of Water Resour Assoc,* 39: 757–769.

Vousta, D., Manoli, E., Samara, C., Sofoniou, M., and Starits, I. (2001).“A study of surface water quality in Macedonia, Greece: speciation of nitrogen and phosphorus.”*Water Air Soil Pollut.,*129(1–4):13–32

Wang, Q., Li, S., Jia, P., Qi, C., and Ding, F. (2013). “A Review of Surface Water Quality Models.” *The Scientific World Journal, 2013(), 1–7.* doi:10.1155/2013/231768.

Water Programme (2007) Global Drinking Water Quality Index Development and Sensitivity Analysis Report. Prepared and published by the United Nations Environment Programme and Global Environment Monitoring System (GEMS)/Water Programme

Wu, T., Qin, B., Brookes, J. D., Shi, K., Zhu, G., Zhu, M., Yan,W.,and Wang, Z. (2015). “The influence of changes in wind patternson the areal extension of surface cyanobacterial bloomsin a large shallow lake in China.”*Science of the TotalEnvironment*, 518–519, 24–30. https://doi.org/10.1016/j.scitotenv.2015.02.090.

Wu, Z., Wang, X., Chen, Y., Cai, Y., and Deng, J. (2018). “Assessingriver water quality using water quality index in Lake TaihuBasin, China.”*Science of the Total Environment*, 612, 914–922. https://doi.org/10.1016/j.scitotenv.2017.08.293.

Young, C. C., Liu, W. C., and Chung, C. E. (2015).“Genetic algorithm and fuzzy neural networks combined with the hydrological modeling system for forecasting watershed runoff discharge.”*Neural Comput. Appl.*26:1631–1643.

Zadeh, L. A. (1965).“Fuzzy Sets.”*Intl J. Information Control,*8:338-353.

Zhang, G., Patuwo, B. E., and Hu, M. Y. (1998). “Forecasting with artificial neural networks: The state of the art.” *Int. Journal of Forecasting*, Elsevier, 14(1), 35–62. doi:10.1016/s0169-2070(97)00044-7

Zhang, X., Srinivasan, R., and Bosch, D.(2009). "Calibration and uncertainty analysis of the SWAT model using Genetic Algorithms and Bayesian Model Averaging."*J. Hydrol*. 374: 307–317, doi:10.1016/j. jhydrol.2009.06.023

Zhang, Z., Zhang, Q., and Singh, V. P. (2018). "Univariate streamflow forecasting using commonly used data-driven models: literature review and case study." Hydrological Sciences Journal, DOI: 10.1080/02626667.2018.1469756

Zhao, Y., Xia, X. H., Yang, Z. F., and Wang, F. (2012). "Assessment of water quality in Baiyangdian Lake using multivariate statistical techniques." *Procedia Environmental Sciences*, 13, 1213–1226. https://doi.org/10.1016/j.proenv.2012.01.115.

Publications

1. Kambalimath, S., Deka, P.C. A basic review of fuzzy logic applications in hydrology and water resources. *Appl Water Sci* 10, 191 (2020). https://doi.org/10.1007/s13201-020-01276-2
2. Kambalimath, S., Deka, P.C. Univariate Daily Streamflow Forecasting using ANFIS and SVM Models - A Case Study on Malaprabha Sub-Basin. *Int. Conf.* on Civil Engineering Trends and Challenges for Sustainability (CTCS 2020).
3. Kambalimath, S. S., Deka, P.C. Performance enhancement of SVM model using discrete wavelet transform for daily streamflow forecasting. *Environ Earth Sci* 80, 101 (2021). https://doi.org/10.1007/s12665-021-09394-z