Study of Time Series Data Mining for the Real Time Hydrological Forecasting: A Review

Satanand Mishra,
WRM &RT Group, CSIR-
Advanced Materials & Processes
Research Institute, Bhopal-
462026, India

C. Saravanan,
Computer Centre,
National Institute of Technology,
Durgapur-713209, India

V. K. Dwivedi,
Dept. of Civil Engineering,
National Institute of Technology,
Durgapur-713209, India.

ABSTRACT

This paper presents a review of runoff forecasting method based on hydrological time series data mining. Researchers are developed models for runoff forecasting using the data mining tools and techniques like regression analysis, clustering, artificial neural network (ANN), and support vector machine (SVM), Genetic Algorithms (GA), fuzzy logic and rough set theories. The scientific community has been trying to find out a better approach to solve the issues of flood problems. Time Series Data mining is paying crucial role for the achieving a real time hydrological forecast. Hydrological Time series is an important class of temporal data objects and it can be find out from water resource management and metrological department. A hydrological time series is a collection of observations of hydro and hydrometeorological parameters chronologically. The wide use of hydrological time series data has initiated a great deal of research and development attempts in the field of data mining. Trend, pattern, simulation, similarity measures indexing, segmentation, visualization and prediction carried out by the researchers with the implicit mining from the historical observed data. The critical reviews of the existing hydrological parameter prediction research are briefly explored to identify the present circumstances in hydrological fields and its concerned issues.

Keywords

Clustering; data mining; runoff; hydrological time series; pattern discovery; regression analysis, ANN, SVM. rough set and fuzzy logic, genetic algorithms.

1. INTRODUCTION

There are large numbers of literature available on Application of Data Mining in Real Time Hydrological Forecasting, watershed management and its impact assessment tools/models. Some of the recent publications which have highlighted hydrological parameter prediction are examined in this paper. In this paper, an introductory background to the hydrology, data mining and its techniques have been described and then a brief insight of literature review on various approaches of clustering, regression analysis, artificial neural network, support vector machines, genetic algorithms, rough sets and fuzzy logics have been discussed[1-13]. Water resource management professional always look for suitable mathematical and computational assessment tools based on minimum but optimum information, cost effectiveness and less time consuming using minimum data with reasonable accuracy. In this regard, several mathematical and computational models for flood forecasting, rainfall and runoff prediction have been developed worldwide in complex and simple approaches. Some conventional popular models namely conceptual models, physically process based models and black box model have been widely used for the hydrological prediction. Large number of data and huge time involvement required for such models and their marginally

superior results compared to the others make them an unfavorable choice. Models like soft lumped computational tools are found user-friendly, as they have been developed based on using minimum data sets, cost effective and less time consuming. But its main faults are that the involves in long calibration and validation process. Data mining techniques such as clustering, regression analysis, artificial neural network, support vector machine, rough sets and fuzzy logic, mathematical simulation are widely used for the hydrological prediction. The hydrologist find regression and artificial neural network models more attractive for the prediction of hydrological forecasting.

Clustering is the process of grouping the data into classes or clusters so that objects within a cluster have high similarity in comparison to one another, but very dissimilar to objects in other clusters. Clustering of time series has received considerable attention in recent years because it is a fundamental task in data mining [6][11-12].

Regression analysis model is developed based on empirical methods used for the modeling and analysis of numerical time series data consisting of values of response variable and of one or more independent variables. Regression Models have been used for prediction including forecasting of time series data, hypothesis testing, modeling of relationships in hydrological parameters and water quality assessment [3][14].

Artificial Neural Networks (ANN) based computing models have been developed for hydrological time series data. ANN is performing like a human brain because it has its own artificial intelligence. The main advantages of the ANN approach over traditional methods is that it needs less data and is capable for long duration forecasting as well as it does not require adequate of particular background science. The main constraint of the model is that it is based on black box because it is generally unable to demonstrate the coefficient of determinants and trend analysis like regression model and model is developed by trial and error approach. Despite of black- box the models have been found more attractive since last one decade and there is a growing trend for use of ANN based hydrological model as it has better performance as compared to conventional models for forecasting [15-19].

Time series data analysis is a new approach along with weighted support vector machine. Typically nonlinear features are introduced through a nonlinear kernel function. While these studies propose a new feature induction algorithm for SVMs. Specifically, new features are induced iteratively. At each step, it weights training examples differently according to the outputs of the current SVM model. The weighted training examples are then used to train a classification model that becomes a new feature for the SVM model [20].

Now a day's comparison of ANN, fuzzy logic and genetic algorithms for rainfall- runoff modeling has carried out. The study described pros and cons of these algorithms and suggest

framework of new algorithm for Rainfall runoff modeling. **(Patil et al, 2012)**

A review of relevant literatures has been done as a part of this work. The focus of the review has been to identify sources of information and techniques that would help for the development of a methodology for the hydrological prediction. This paper gives an overview of previous work done on related topic that provides the necessary background for the purpose of this research. The literature review concentrates on runoff prediction. Many models have been developed for forecasting the runoff. In spite of substantial efforts by hydrologist/scientist cannot yet make any accurate models for runoff predictions for the river Brahmaputra. The hydrological prediction literature contains many comparisons and evaluations of prediction algorithms. There are several literature reviews that focus on evaluation of rainfall, runoff and water level prediction algorithms based on gauge or sensor data and other measurements.

2. REVIEW OF CLUSTERING APPROACH FOR THE HYDROLOGICAL STUDY

Hydrological forecasting is an important aspect of hydrological services on economy and society, especially making prediction of disastrous hydrological phenomena. Currently, the main applications of cluster algorithms in hydrological forecasting include rainfall monitoring, flood forecasting, and water quality monitoring, inflow and outflow calculation. The main clustering methods contain k-means method, hierarchical method, Similarity search, SOM method and FCM method **(Tian et al., 2014)**. **Purviya R. Et al. (2014)** described clustering can be done by the different number of algorithms such as hierarchical, partitioning, grid and density based algorithms. This paper is original concerns in two main aspects. First, it provides an evolutionary algorithm for clustering starting from data mining mechanism, tasks and its learning. Second, it provides a taxonomy that highlights some very important aspects in the context of clustering algorithms, namely, hierarchical, partitional algorithms, density based, grid based and model-based. Developing a hydrological model based on past records is crucial and effective in many water resources applications such as optimal reservoir operation, drought management, flood control, hydropower generation and sustainable development of watershed area, etc[13]. **Ramoni et al. (2000** presented a study on BCD, a Bayesian algorithm for clustering by dynamics. BCD is basically an unsupervised algorithms based on agglomerative clustering method. The clustering result is evaluated mainly by a measure of the loss of data information induced by clustering, which is specific to the proposed clustering method. They also presented a Bayesian clustering algorithm for multivariate time series [21]. The algorithm searches for the most probable set of clusters given the data using a similarity-based heuristic search method. The measure of similarity is an average of the Kullback–Liebler distances between comparable transition probability tables. **Van Wijk and Van Selow (1999)** analysed an agglomerative hierarchical clustering of daily power consumption data based on the root mean square distance. How the clusters distributed over the week and over the year were also explored with calendar-based visualization [22]. **Vlachos et al. (2003)** introducing a novel anytime version of k-Means clustering algorithm for time series. In this work we address these challenges by introducing a novel anytime version of k-Means clustering algorithm for time series. It is an approach to perform incremental clustering of time-series at various resolutions using the Haar wavelet transform. Using *k-Means*

clustering algorithm, for the next level of resolution, they modified the final centers at the end of each resolution as the initial centres [23]. **Li and Biswas (1999)** described a clustering methodology for temporal data using the hidden Markov model representation. The proposed continuous HMM clustering method can be summarized in terms of four levels of nested searches. The HMM refinement procedure for the third-level search starts with an initial model configuration and incrementally grows or shrinks the model through HMM state splitting and merging operations. They generated an artificial data set from three random generative models: one with three states, one with four states, and one with five states, and showed that their method could reconstruct the HMM with the correct model size and near perfect model parameter values.[24]. **Wei T. et al., (2014)** described **c**luster analysis techniques in data mining play an important role in the study of meteorological applications. **Warren Liao T., (2005)** presented the basics of time series clustering are presented, including general-purpose clustering algorithms commonly used in time series clustering studies, the criteria for evaluating the performance of the clustering results, and the measures to determine the similarity/dissimilarity between two time series being compared, either in the forms of raw data, extracted features, or some model parameters. There seems to be an increased interest in time series clustering as part of the effort in temporal data mining research. Clustering is necessary when no labeled data are available regardless of whether the data are binary, categorical, numerical, interval, ordinal, relational, textual, spatial, temporal, spatio-temporal, image, multimedia, or mixtures of the above data types **[25].**

Uncertainties in climate change and human activities have great impacts on water resources. It mainly includes rainfall change, runoff decrease, flood disaster and so on. Thus, the reasonable configuration and integrated planning of water resources are important parts of the water resources management. Clustering algorithms classify different watersheds with similar hydrological characteristics. Changes of precipitation, basin evaporation and snow density under certain climatic conditions, have been studied. At present, the clustering methods in this field contain the k-means, hierarchical method, SOM method and FCM method.

In view of the clustering algorithms, k-means and hierarchical methods are widely used in hydrological applications mentioned above for their typicality and distinct advantages.

3. REVIEW OF RAINFALL –RUNOFF MODELING BASED ON REGRESSION APPROACH

Adequate literature is available on stochastic modeling. Generally, the rainfall-runoff forecasting model needs to fulfil four principles: parsimony, modesty, accuracy and testability (Abrahart and See, 2000). It is obvious that a large number of models have been developed and used both as research and as operational tools **[26].** The most widely-applied procedure for hydrological regionalization applied to date has been the so-called Method of Residuals, in which the classification and modeling are carried out simultaneously, with the appropriate models being developed by application of MLRA. The hydrological index variable (quantile) is regressed upon catchment and rainfall characteristics for the whole data set. The residuals, i.e. the differences between the observed and computed values of the index variable, are plotted geographically in order to identify groups of these differences that are similar in both magnitude and sign and can therefore be regarded as a sub-region; and the regression analysis is repeated

for the sub-regions identified and then generalized across the whole catchments[27]. **P. P. Mujumdar & D. Nagesh Kumar (1991)** investigated ten candidate models of the Auto-Regressive Moving Average (ARMA) family for representing and forecasting monthly and ten-day streamflow in three Indian rivers namely Cauvery, Hemavathy and Malaprabha. They used Minimum Mean Square Error (MMSE) and Maximum Likelihood (ML) criteria for selection of best models for forecasting and representation of data respectively. The validation criteria of selected are based on the significance of the residual mean, significance of the periodicities in the residuals and significance of the correlation in the residuals. Based on the ML criterion ARMA(4,0), ARMA(2,1) and ARMA(3,1) are developed for the synthetic generation of the three monthly series of the Rivers Cauvery, Hemavathy and Malaprabha, respectively. While the ARMA (4,0) model is selected for the ten-day series of the Malaprabha River. The MSE value is minimum of ARMA (1, 0) model in all the cases studied and is recommended for use in forecasting flows one time step ahead. Validations tests revealed that the models selected by the two criteria pass all the tests and hence these models are recommended for use in practice for the three rivers [28].

Wang, Wen-Chuan, et al (2009) autoregressive moving-average (ARMA) models, artificial neural networks (ANNs) approaches, adaptive neural-based fuzzy inference system (ANFIS) techniques, genetic programming (GP) models and support vector machine (SVM) method are presented by authors to examined the long-term observations of monthly river flow discharges [29]. **P.E. Naill and M. Momani (2009)** used Box-Jenkins methodology to build ARIMA model for monthly rainfall data taken for Amman airport station for the period from 1922-1999 with a total of 936 readings. In their research, ARIMA (1, 0, 0) (0, 1,1) 12 model was developed. This model is used to forecasting the monthly rainfall for the upcoming 10 years to help decision makers establish priorities in terms of water demand management [30]. **Kothyari et al. (1997)** studied rainfall and temperature (i.e., long-term monsoon rainfall, number of rainy days during the monsoon season, and annual maximum temperature) from three stations at Agra, Dehradun and Delhi for evaluating the changes in regimes in the upper and middle parts of the Ganga basin in northern India [31]. **Makridakis S. and Hison M.** presented various aspects of the Box-Jenkins methodology to ARMA models in this paper. The major concern of this study is to identify the post sample forecasting methods and improve the accuracy in order to ascertain static form of data[30]. **Boudaghpour, S. et al (2014)** explain the stochastic AR (1) and AR (2) models which used for the case study of TSS in Lighvan Chia basin. They further coupled four autoregressive moving average models including, ARMA (1,1), ARMA (2,2), ARMA (1,2) and ARMA (2,1). The results of AR (1) and AR (2) models showed almost a perfect match between measured and the calculated values of TSS among these models. As a results, the AR (1) models are introduced to assess and predict the volume of TSS and sediments with satisfactory achievements [32]. **Tesfaye, Y. G. et al. (2006)** introduced the model identification and simulation techniques which have developed based on a periodic autoregressive moving average (PARMA). The purpose is to capture the seasonal variations in river flow statistics. The innovations algorithm is used to obtain parameter estimates. The model is then tested to monthly flow data for the Fraser River in British Columbia .A cautious look over statistical analysis of the PARMA model residuals, including a truncated Pareto model for the extreme tails, generate a realistic simulation of these river flows [33]. **Musa, J. J., (2013)**, a stochastic approach is

presented by an Autoregressive Moving Average model (ARMA). Autoregressive (AR) model, Autoregressive Moving Average (ARMA) model and Autoregressive Integrated Moving Average (ARIMA) model are used to analyse 22 years (1990-2011) observations of flow discharge from the Meteorological Station at Shiroro [34]. **Toth, E. et al. (2000)**, in their study only rainfall depths is used as only input data by which they tried to compares the accuracy of the short-term rainfall forecasts obtained with time-series analysis techniques. Linear stochastic auto-regressive moving-average (ARMA) models, artificial neural networks (ANN) and the non-parametric nearest-neighbours method techniques are used for developing model on the Apennines Mountains, Italy. The study analyses and compares the relative advantages and extent of each time-series analysis technique, used for predicting rainfall forecasts for lead-times varying from 1 to 6 h. The results indicate ANN techniques are mere accurate than the ARMA and nearest-neighbours method [35]. **Chakraborty, S. et al., (2013)** conducted a study to develop a stochastic time series model, capable of prediction of rainfall and runoff. The watershed they had been chosen for study was Kelo watershed of Mahanadi river basin of Chhattisgarh state, India. Six years data of rainfall and runoff of the watershed was collected and used for the development of model. They were tried Autoregressive (AR) models of order 0,1 and 2. The goodness of fit and adequacy of models were tested by Box- Pierce Portmanteau test, Akaike information Criterion and by comparison of historical and generated data correlogram [36].

4. REVIEW OF RAINFALL –RUNOFF MODELING BASED ON ANN APPROACH

The artificial neural network (ANN's) models are powerful prediction tools for the relation between rainfall and runoff parameters. The results will support decision making in the area of water resources planning and management. The use ANN's of is becoming increasingly common in analysis of hydrological and water resources problem. The ANN was developed and used to model the rainfall-runoff relationship for various catchments. It is also helpful in urban planners and managers undertake the necessary measures to face the bad predictions. It has great advantage to avoid losses in public and private properties, health and ecological hazards that are likely to occur due to floods. The ANN models have been used in various aspects of science and engineering applications because of its ability to model both linear and nonlinear systems without the need to make any assumptions like most traditional statistical approaches. The ANN is widely used in river flow prediction, river flood prediction, water quality parameter prediction, rainfall-runoff process, evaporation etc time series prediction. For the hydrological applications, a multilayer feed-forward backpropagation algorithms is used. It is a set of a large number of interconnected nodes, arranged in an input layer, an output layer and one or more hidden layers. The transfer function selected for the network was the sigmoid function [4].
In this section the potentially very powerful data mining tool artificial neural network and its various approaches discussed.

4.1 Neural network based on Back Propagation Approach

Back propagation is a common method to train an Artificial Neural Network. The objective of training is to find the set of weights between neurons that determine the global minimum of error function. The Back propagation algorithm is used in layered feed forward ANNs. Signals are forwarded from one layer to other layer and errors are propagated backwards The

main idea of the Back propagation algorithm is to minimize this error, until the ANN learns the training data.

Abhishek et. al. (2012) estimated average rainfall using back propagation algorithm along with the multilayered artificial neural network. By using artificial neural network model two phases of learning cycle, one to propagate the input pattern through the network and other to adopt the output by enhancing the weights in the network have been analyzed. It is found that as the number of neurons increases in an ANN, the MSE decreases and BPA is the best algorithm out of the three tested **[15]. Adamowski, J. and Karapataki, C. (2010)** compared three different type artificial neural network and multiple linear regression model for peak weekly water demand forecasting. 20 Levenberg-Marquardt artificial neural network models, 20 resilient back-propagation ANN models, and 20 conjugate gradient Powell Beale ANN, 20 multiple linear regression models, models were developed and their relative performance was compared .This study indicated ANN method was a more accurate prediction method with respect to other conventional data mining approaches**[17]. Adamowski (2008)** compared multiple linear regression, time series analysis and artificial neural networks as techniques for peak daily water demand forecast modelling. **[18]. Agarwal, A., and Singh, R. D. (2004)** have developed a Multi-layer back propagation artificial neural network (BPANN) model to simulate rainfall-runoff process **[31].** An emerging field of research characterized by a wide variety of techniques for the application of rainfall-runoff modelling and flood forecasting using ANN has been discussed. This review underline the need for clear guidance in current modelling practice, as well as the analogy of ANN methods with more conventional statistical models **[37].** A depth discussion of the development and application of artificial neural networks to flow forecasting in two flood prone UK catchments using real hydrometric data carried out. The results obtained for validation forecasts were of comparable quality to those obtained from operational systems for the River Amber. The strength of the ANN to adjust with missing/nullified data and to "learn" from the event currently being forecast in real time makes it an appealing alternative to conventional flood forecasting models **[38]. Devi, C. J. et. al. (2012)** present a neural network-based algorithm for predicting the temperature. **[37]. Edossa and Babel** have been modelled a nonlinear stream flow forecasting modelling technique along with ANN for study the drought forecasting. In this available dataset was divided into two independent sets using a split sampling tool of neural network software. The stream flow predicted using the model on weekly time step compared with the major stream flow during the validation period **[39]. Hsieh et. al.** investigated the prediction system with single-point river stage, upstream-downstream river flow forecasting, and rainfall-runoff hydrological process **[40]. Kumar and Yadav (2013)** have been introduced ANN as one of most important modern tool used for data processing and hydrological forecasting. Feed forward neural network trained with Levenberg-Marquardt back propagation algorithm for rainfall forecasting. The regression coefficient was measured for accuracy of developed model **[41].**

Kumarasiri, A. D., and Sonnadara, U. J. (2008) have been attempted to implement both short term and long term forecasting with the help of feed forward back propagation artificial neural network for daily occurrence and annual depth of rainfall at a single meteorological station. Two neural network models were developed and trained with back propagation algorithm for one day ahead and one year ahead prediction **[42]. Luc, K. C. et.al. (2001)** presented the results of a study investigating the application of ANNs to forecast the spatial distribution of rainfall for an urban catchment. Artificial

neural network technique is used for rainfall forecasting. Three different type of ANNs (MLFN, PRNN, TDNN) were developed and their performance are compared. All three networks had comparable performance when they were developed and trained to reach their optimal complexities **[43].**

This research evaluated the capacity of ANNs to model with accuracy the monthly rainfall-runoff process **[44].** A survey of 43 papers handling with the implementation of neural network models for the forecasting and prediction of water resources variables is undertaken with respect of the modelling process adopted**[45]. Minns, A.W. and Hall, M.J.** developed a ANN model for rainfall runoff forecasting. ANN model is trained with back propagation algorithm. The application of different standardization factors to both training and verification sequences have underlined the importance of such factors to network performance **[46]. Mittal et. al. (2012)** proposed a Dual Artificial Neural Network, to improve the model performance in terms of extreme values. Proposed model performance compared with feed forward ANN and extracted result indicate that D-ANN model perform better than FF-ANN model **[19]. Mutlu et. al. (2008)** presented an evolution to use of ANN models to forecast daily flows at multiple gauging station in Eucha watershed. In this study the MLP and radial basis feed forward neural network (RBFNN) were developed and their abilities to predict stream flow were compared **[47].** The performance analysis of the neural network is compared with the statistical technique of synthetic inflow generation **[48]. Riadet.et. al**. (2004) developed an ANN model to analyze the rainfall runoff relationship. Multilayer perceptron neural network based model is more confined to predict river runoff than classical regression model **[49]. Sarkar A. and Kumar R. (2012)** examined the study to model the event based rainfall runoff process. A case study has been done to develop event based rainfall runoff model for the basin to simulate the runoff. Obtained result demonstrates that ANN model are able to provide a good simulation for event based forecasting **[50].** A performance comparison of three artificial neural network techniques for adding simulated river flows, based on three different neural network structures has been done. These network structures are: the radial basis function neural network (RBFNN), the simple neural network (SNN), and the multilayer perceptron neural network (MLPNN)**[51]. Sharda et. al. (2006)** analyzed three hilly micro watersheds to assess the impact of morphological characteristics on water flow. To predict runoff and base flow, a artificial intelligence model is developed using multivariate adaptive regression splines method. The predicted value from proposed model found good and it is observed that MARS model have good capability to simulate runoff in hilly area **[52].** A wide variety of standard statistical performance evaluation measures were employed to evaluate the performance of various ANN models **[53]. Wu et. (2005)** demonstrated an application of ANN for runoff and stream flow forecast. **[54].**

4.2 Regression Based Neural network Approach

The general regression neural network (GRNN) is a one-pass learning algorithm with a highly parallel structure **[52].** The GRNN is used for estimation of continuous variables, as in standard regression techniques. It is related to the radial basis function network and is based on a standard statistical technique called kernel regression. The algorithmic form can be used for any regression problem in which an assumption of linearity is not justified. **Cigizoglu, H. K. (2005)** employed an ANN algorithm, generalized regression neural network (GRNN), in monthly mean flow forecasting**[55]. Cigizoglu, H. K. and Alp,**

M. (2006) deployed generalized regression neural network (GRNN) for river suspended sediment estimation **[56]**. **Dreiseitl, S., and Ohno-Machado, L., (2002)** encapsulate the similarities and difference between logistic regressions, artificial neural network and compared their performance with other conventional machine learning algorithm. On the basis of comparison ANN and logistic regression found better than the conventional machine learning algorithm until newer machine learning algorithms, such as support vector machines and other kernel-based algorithms, can prove to be significant **[38]**. **Koutsoyiannis, D. (2007)** discussed the research paper, generalized regression neural network for evapotranspiration modeling**[57]**. **Kisi, O. (2006)** investigated the ability of generalized regression neural networks technique in evapotranspiration (ET0) modelling obtained using the FAO Penman-Monteith (PM) equation. The performance of GRNN model was compared to the empirical model using mean square error, mean absolute error. Result obtained based on comparison shows that GRNN model strongly employed in evapotranspiration modelling **[58]**. **Specht, D. F.** describes a new network that provides estimates of continuous variables and converges to the linear or nonlinear regression surface. The proposed memory based general regression neural network (GRNN) features fast learning that does not require an iterative procedure, and a highly parallel structure. It can be used for prediction, modelling, mapping, and interpolating or as a controller. The algorithmic form can be used for any regression problem in which an assumption of linearity is not justified. **Wang, Z. L. and Sheng, H. H. (2010)** introduces the application of generalized regression neural network (GRNN) model to forecast annual rainfall. The result of proposed method is compared with the back propagation neural network and regression analysis method. Compared with the traditional linear model and BP neural networks, the GRNN has smaller prediction error. The results and accuracy of GRNN model comes better than that of BP neural network for annual rainfall **[59]**.

4.3 Wavelet Neural Network Based Approach

In this section, presented a neural network technique combined with wavelet analysis. Wavelet analysis is employed to pre-process the data **[60]**. The wavelet decomposition of non-stationary time series into different scales provides an interpretation of the series structure and extracts the significant information about its history, using few coefficients. For these reasons, this technique is largely applied to times series analysis of non-stationary signals. **Adamowski, J. and Sun, K. (2010)** proposed a method, for flow forecasting applications in non-perennial rivers in semi-arid watersheds which is based on coupling discrete wavelet transforms (WA) and artificial neural networks (ANN) techniques**[61]**. **Adamowski, J. et. al.(2012)** developed three model; multivariate adaptive regression spline (MARS), wavelet transform artificial neural network (WA-ANN), and regular artificial neural network (ANN) for runoff forecasting applications **[62]**.

Adamouski, J. and Chan, H. F. (2011) described a new approach based on pairing discrete wavelet transforms (WA) and artificial neural networks (ANN) for groundwater level forecasting[8]. **Adamowski and Belayneh (2012)** compared the effectiveness of three data-driven models for forecasting drought conditions in the Awash River Basin of Ethiopia **[64]**. **Campisi-Pinto, S. et. al. (2012)** introduced a new hybrid approach which is the combination of wavelet-denoising and back-propagation artificial neural networks (ANNs) for urban water demand forecasting. The proposed model performance

were compared with non-coupled ANN and Linear Multiple Regression models**[65]**. **Cannas B. et. al. (2005)** applied neural network and wavelet analysis to predict the hydrologic behaviour of the runoff. In this research paper data is pre-processed for neural network input and output using continuous and discrete wavelet transforms. Proposed approach is compared with ANN. comparison result showed that the networks trained with pre-processed data perform better than the networks trained with undecomposed noisy raw signals **[60]**.

Partal, T. and Cigizoglu, H.K. (2009) proposed a wavelet–neural network method to predict the daily precipitation from meteorological data. wavelet neural network is a integtration of discrete wavelet transform (DWT) and artificial neural networks (ANN) **[66]**.

Ramana, R. V. et. al. (2013) combined the wavelet technique with Artificial Neural Network (ANN) for rainfall prediction **[67]**. **Santos and Silva (2012)** proposed a new wavelet and artificial neural network hybrid model for daily stream flow forecasting**[68]**. **Wei, S. et. al. (2013)** introduced a wavelet-neural-network (WNN) hybrid modelling approach for monthly river flow evaluation. This proposed approach combines discrete wavelet multi-resolution decomposition and a back propagation (BP) feed-forward multilayer perceptron (FFML) artificial neural network (ANN). The Bayesian regularization and Levenberg-Marquardt algorithm were use to perform the network modelling. The results from WNN hybrid model was compared with single ANN model. The comparison shows that WNN hybrid model perform better than the single ANN model **[69]**.

5. REVIEW OF RAINFALL –RUNOFF MODELING BASED ON A SUPPORT VECTOR APPROACH

Rainfall-runoff simulation and prediction in watersheds is one of the most important tasks in water resources management. In this research, an adaptive data analysis methodology, ensemble empirical mode decomposition (EEMD), is presented for decomposing annual rainfall series in a rainfall-runoff model based on a support vector machine (SVM) **[104]**.

5.1 Support Vector Machine in Linear Regression

Support vector machine is based on the theory of statistical learning and structural risk minimization. In this learning by the help of support vectors a linear discriminant function separating two classes has been built **[38]**. For nonlinear class of problem, either penalty for misclassification can be introduced or nonlinear mapping into a space of higher dimension can be performed by the help of different Kernel functions **[70]**. Vapnik theory plays a significant role in generalized implementation of support vector machine and for the statistical characteristics of generated input and formulated machine learning model. The potentially very powerful data mining tool as a support vector machine and its various approaches discussed here. Some introductory background of support vector machines based on least square and hybrid model based support vector machine approaches for the hydrological parameter prediction is carried out.

5.2 Support Vector Machine Based Approaches

Support vector machine (SVM) was first practically come into existence in 1992, by Boser, Guyon, and Vapnik in COLT-92. SVMs are a set of uniquely related supervised learning concepts

used for classification and regression task[71]. This theory relates it to a family of generalized linear classifiers. In other way, Support vector machine (SVM) is a binary classification and simple regression data mining tool that uses machine learning concept to maximize forecasting accuracy while automatically avoiding local minima and over fitting to the data. Support Vector Machine can be summarized as a system which use feature space hypothesis concept of a linear functions in a d-dimensional space, simulated and trained with a learning algorithm generated from statistical learning theory. There is vast mining tools are available for data analysis, clustering, regression and pattern discovery. One of the tool is Support vector machine so a guide for users for the better understand ability of Support Vector Machines as a widely used classifier and describe the effect of the SVM parameters on the resulting classifier, how to select good values for those parameters, data normalization, factors that affect training time, and software for training SVMs. Author provides the user with a basic understanding of the theory behind SVMs and focus on their use in practice [72]. To familiarize the concept of Support Vector Machines for data mining this paper introduce support vector machine and highlight the advantages thereof over existing data analysis techniques [73]. The modeling of the relationship between rainfall and river discharges of the Fuji River using the SVRBFN is presented. The main advantage of this approach is that the structure of the network can be obtained objectively, as the SVs of the SVM are obtained from the constrained optimization for a given error bound [74]. Support vector regression is a new regression procedure in water resources for predicting suspended sediments load in rivers [75].

A novel regression approach, termed as the Rough Margin Support Vector Regression (RMSVR) network, is the approach to adopt the concept of rough sets to construct the model obtained by SVR and fine tune it with a robust learning algorithm. Simulation results of the proposed approach have shown the effectiveness of the approximated function in discriminating against outliers [76]. The purpose of the study of the SVM, is to develop a parsimonious model used little operation gage that accurately simulates semi-arid regions by using SVM models [77]. Recommended learning in hydrological prediction environment are, firstly, to examine the limitations in applying conventional method for evaluating the data driven forecasting model performance, and, secondly, to present the new performance evaluation methods that can be used to evaluate hydrological data-driven forecasting model with consistency and objectively. The Relative Correlation Coefficient (RCC) is used to estimate the forecasting efficiency relative to the naïve model in data driven forecasting [78]. Support vector machine is a useful tool for knowledgeable data discovery. A novel clustering based Short Term Load Forecasting (STLF) using support vector machine (SVM). The forecasting is performed for the 48 half hourly loads of the next day. The data considered for forecasting contains 2 years of half hourly daily load and daily average temperature. The proposed architecture is implemented in Matlab[79]. Time series data analysis is a new approach along with weighted support vector machine. Typically nonlinear features are introduced through a nonlinear kernel function. While these studies propose a new feature induction algorithm for SVMs. Specifically, new features are induced iteratively. At each step, it weights training examples differently according to the outputs of the current SVM model. The weighted training examples are then used to train a classification model that becomes a new feature for the SVM model [20]. The accuracy of support vector machines (SVM), which are regression procedures, in modelling reference to evapotranspiration (ET0). Obtained result comparison reveals that the support vector machines could be employed successfully in modelling the ET0 process [80]. In this study, four simple dynamic methods and two supervised learning techniques including a linear regression model, a quadratic regression model, an original grey prediction model, a back-propagation neural network model, and an epsilon-SVM regression model were investigated for the forecasting of flood stage one hour ahead for early warning of flooding hazards [81].

Li H.P.(2010) et al introduced a modified SVM based prediction framework which improves the predictability of the inflow using climate data from the prior period [14]. **Lingras and Butz (2004)** presented unique classifier techniques with support vector machine and rough set theory [82]. **Liong Yui-Shie and S.Chandrasekaran (2002)** study reveals a novel machine learning and regression technique and application of Support Vector Machine as a robust forecasting tool has been shown with its implementation in the Bangladesh flood stage data [78]. **Lin Y.J. et al (2006)** presented SVM as a promising method for hydrological prediction. The SVM prediction model is tested using the long term observation of discharges of monthly river flow discharges in the Manwan Hydropower Scheme. Obtained result demonstrates SVM as a very potential candidate for the long term discharges [79]. **Mamat M and Samad A.S.** compares the performance of Radial Basis Function and Support Vector Regression in time series forecasting. Both methods were trained to produce one step ahead forecasting on two chaotic time series data: Mackey Glass and Set A data. [83]. **Misra et al (2009)** demonstrated to use Support Vector Machines (SVM) to simulate runoff and sediment yield from watersheds. They have simulated daily, weekly, and monthly runoff and sediment yield from an Indian watershed, with monsoon period data, using SVM, a relatively new pattern-recognition algorithm [84]. **Mountrakis et al (2010)** presented a review over remote sensing implementations of support vector machines[85]. **Solomatine P.D. (2002)** described the application of the artificial neural networks, fuzzy rule-based systems, M5 model trees, support vector machine, chaos theory to the problems of water management and control in this research [86]. **Shahbazi and Pilpayeh (2012)** have implemented the concept support vector machine for forecasting of river flow downstream through examining the data collected from upstream. Results have shown that SVM acts well and efficient in preserving the monthly and yearly features [87]. **Torabi M. and Hashemi S. (2012)** presented an approach to forecast three-day ahead hourly electric energy consumption. To exact energy consumption patterns, Neural Network and Support Vector Machine are adapted in a novel manner [88]. **Wu et al (2004)** implement concept of Support Vector Regression (SVR) for the use of identifying prediction of travel time and compare its result to other travel time forecasting method using real highway traffic datasets [89]. **Yang et al (2009)** used the concept of non- stationary and volatile data with localized support vector regression model [90]. **Yi D. and Wei C.** experiment shows that the fitting value which obtains using the return to first would be more precise than directly[91]. **Xian and Zeng** presented a two-phase, efficient, and fair evaluation method for DRMs (digital right management system) basing on SVM. Influence of three difference methods and test set number on evaluation result id discussed [92]. **Yoon H. et al** developed two nonlinear time series models for predicting ground water level fluctuations using artificial neural networks (ANNs) and support vector machines (SVMs). The results of the model performance show that root mean square value of ANN models is lower than those of SVM in model training and testing stages [93].

5.3 Least Square Support Vector Machine Approach

Classification problems have arisen in many applications, attracting many researches to develop advanced classifier techniques. A method called Least Square Support Vector Machines (LS-SVM) for pattern recognition and function estimation has been introduced in the framework of statistical learning theory. Since then there is a growing interest on this kernel method for its interesting features. In this section, a least squares version (LS-SVM) is explained. LS-SVM expresses the training in terms of solving a set of linear equations instead of quadratic programming as for the standard SVM case. Iterative training algorithm for LS-SVM based on a conjugate gradient method is then applied. **Bhagwat P. and Maity R (2013)** have been explored potential of least square support vector regression model in the context of streamflow prediction using Hydroclimatic inputs**[72]. Espinoza M.** presented a primal dual formulation of Least Square Support Vector Machine**[74]. Hwang H.S. et al (2012)** presents a Least Square Support vector machine (LS-SVM) approach for forecasting nonlinear hydrological time series. This paper is to examine the feasibility using LS-SVM in the forecasting of nonlinear hydrological time series by comparing it with a statistical method such as Multiple Linear Regression (MLR) and a heuristic method such as a Neural Network using Back-Propagation (NNBP). In the experimental results, LS-SVM showed superior forecasting accuracies and performances to those of MLR and NNBP **[94]. Samsudin et al (2011)** projected a new, unique and novel hybrid forecasting model known as GLSSVM, which merges the group technique of data handling and least square support vector machine**[95]. Shabri A. and Suhartono (2012)** investigate the ability of least square support vector machine (LSSVM) model to enhance the accuracy of prediction. Grid search and cross-validation techniques are used to determine the essential LSSVM parameters in the forecasting processes**[96]. Shah Shiloh R.(2005)** described the concept and characteristics of least squares support vector machines and then produced detailed result using Least Square Support Vector Regression which shows significant improvement over Support Vector Machine (SVM) **[97]. Suykens et al (2001)** introduced a unique method which can overcome the drawbacks of sparseness and the estimation of the support vectors values. They also introduce a sparse approximation procedure for weighted and un-weighted version of LS-SVM. The result showed how to obtain robust estimates within the LS-SVM framework in case of outliers and heavy tailed non-Gaussian error distributions **[98]**.

5.4 Hybrid Model Approach

A new comprehensive and more refine approach where machine learning technique and data mining technique are combined together are termed as Hybrid model approach for hydrological prediction. Classification and regression analysis have been performed in an efficient way, leads to a accurate result simulation. In this section, hybrid approach is studied through a various research papers.

In this study **Botsis D. et al (2011)** presented a performance comparison between Support vector regression and multilayer feed-forward neural network models with respect to their forecasting capabilities **[71]. Flint and Flint (2012)** proposed an evaluation on the impact of climate change on water resources and scientific components of landscape is a necessary part of this investigation**[75]. Ismail et al (2012)** proposed a innovative hybrid approach based on a combination of two approaches: Self Organizing Map (SOM) and Least Square Support Vector Model (LSSVM), termed as a SOM-LSSVM model for river flow forecasting [99]. **Lin F. et al (2013)**

propose a hybrid model approach which joins locally linear embedding algorithm and support vector machine to predict the firms which based on past financial performance of data sets**[100]. M. Abbasi (2013)** Study proposed a novel method of forecasting municipal solid waste (MSW) generation **[20]. Ma Xixia et al (2011)** presented a combined hybrid model of chaos theory; wavelet and support vector machine was built to overcome the limitations including challenges in determination of order of nonlinear models and low prediction accuracy which the simulated accuracy is high in runoff forecasting **[80]. Maity R. and Khasid S.S. (2009)** presented an approach for monthly streamflow prediction. In this paper using the concept of hydroclimatological association rainfall-runoff relationship over a catchment is explained **[81][84]. Mohammadi K.et al (2006)** tried to achieve a goal to minimize the error of specific season of the year as well as for the complete series. Goal programming was used to estimate the ARMA model parameters **[14]. Pandhiani and Shabri (2013)** explore the least square support vector and wavelet technique (WLSSVM) in the monthly stream flow forecasting. This is the new hybrid technique. The Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Correlation(R) statistics are used for evaluating the accuracy of the WLSSVM and WR models**[101]**.

Sivakumar et al (2000) employed the correlation dimension method, provided preliminary evidence of the existence of Chaos in the monthly rainfall-runoff process at the Gota basin in Sweden **[102]. Wang et al (2010)** presented a SVM model with chaotic genetic algorithm (CGA) as a promising method for hydrological prediction**[103]. Wang W.C. et al (2013)** described an adaptive data analysis methodology, ensembles empirical mode decomposition (EEMD), for decomposing annual rainfall series in a rainfall-runoff model based on a support vector machine (SVM) **[104]. Wei L.H. and Billings A.S. (2006)** proposed the long term prediction of non-linear dynamical time series, based on multiresolution wavelet models, from historically observed data sets is investigated and directed prediction approach is introduced **[105]. Yu X. (2004)** demonstrates a combined application of chaos theory and support vector machine (SVM) in the analysis of chaotic time series with a very large sample data record the various parameters inherent in chaos technique and SVM are optimized, with the assistance of an evolutionary algorithm, to yield the minimum prediction error **[106]**.

6. REVIEW OF RAINFALL –RUNOFF MODELING BASED ON GENETIC ALGORITHMS APPROACH

Genetic algorithm (GA) is a new computing paradigm, inspired by Darwin's theory of evolution. Individual's population, each exploring a possible solution to a task which is initially created at random. After that pairs of individuals associated (crossover) to generate offspring for the next generation (**Pujari, 2006).** Genetic algorithm is an adaptive global search method that parodies the metaphor of natural biological evolution. Based on Darwin's theory of evolution, the better sub generation in GA will survive and generate the next generation. Naturally, the best generation will have better presentation to get with the conditions. The method can be applied to an extremely wide range of optimization problems. The genetic algorithm differs from other search methods in that this algorithm searches among a population of points, and works with a coding of the parameter set, rather than the parameter values themselves. It also uses objective function information without any gradient information. Owing to its ability to achieve the global or near global optimum, this algorithm has been applied to a large number of combinatorial optimization problems. The standard

genetic algorithm can be defined based on the following equations:

$$SGA = (C, E, P_0, M, \varphi, \Gamma, \emptyset, T)$$

Where C is the initial population coding scheme; E is the fitness function; P_0 is the initial population; M is the scale of population; φ is the selection operator; Γ is the crossover operator; $\emptyset$ is the mutation operator; T is the stop criterion. (**Mutao H, 2007**)**[104]**.

GA operates on a population of potential solutions by applying the principle of survival of the fittest to achieve an optimal solution. Every solution in the temporary population is ranked against other solutions based on a fitness criterion. **Li S et al. (2009)** proposed an improved genetic algorithm to solve the disadvantage of the standard genetic algorithms inaccuracy of coding and low-precision. The improved genetic algorithm gets better results in calibration of the three-source XinAnJiang model **[108]. Patil et al. (2012)** compared the ANN, fuzzy logic and genetic algorithms for rainfall- runoff modeling **[109]. Banik et al. (2009)** investigated the development of an efficient model based on Fuzzy logic System for the prediction of monthly monsoon rainfall for a number of stations **[110].**

7. REVIEW OF RAINFALL –RUNOFF MODELING BASED ON ROUGH SET THEORY AND FUZZY LOGIC APPROACH

Rough set theory is a mathematical theory to deal with uncertainty and incompleteness of data which was proposed by Pawlak. It is based on the capability of classifying the data observed and measured directly. The Knowledge is based on the ability of classifying objects. For the analysis of knowledge dependency none of other additional information is needed like probability in statics and weight in fuzzy set. On the basis of Rough set theory, single element medium and long term classification forecast model has developed using historical hydrological data. Medium and long term hydrological forecasting are crucial to the optimal distribution and utilization of water resources. The main method for medium and long term forecasting is multi factor forecasts and single factor forecasts.

Dong et al. (2004) developed rough set theory based model which reflects the non linear relations between the input and output effectively. The model composed of the decision rules which are very simple and could be comprehended easily and the model is more credible if the quantity of swatches is larger **[111]. Hundecha et al. (2001)** described Fuzzy logic model which is based on the theory of fuzzy sets. In a fuzzy set the boundary is not clearly defined and partial membership of an element is possible. Each element of the set is assigned a membership value which can be between 0 and 1 inclusively. The function that assigned this value is pointed to as the membership function related with the fuzzy set. Fuzzy numbers are special type of fuzzy sets defined on the set of the real numbers. Fuzzy numbers are generally defined by using membership function which have triangular shapes and are given as $(a_1, a_2, a_3)_T$ such that $a_1 \leq a_2 \leq a_3$, are trapezoidal shapes that are expressed as $(a_1, a_2, a_3, a_4)_R$ such that $a_1 \leq a_2 \leq a_3 \leq a_4$ **[77]. Mahabir et al. (2003) proposed a** Fuzzy logic approach could be used to simulate actual component of hydrologic processes such as snowmelt, runoff, evaporation and basin response. On the availability of vast data the fuzzy logic model physically developed. The main objective of this study was to develop the potential viability of an alternate approach for water supply forecasting based on fuzzy logic. Based on fuzzy set theory,

fuzzy logic is being applied in several fields where empirical relationships are not broadly defined **[112]. Alvisi et al. (2006)** presented data driven water level forecasting models on the basis of fuzzy logic and ANN. The analyses are made with great attention to the reliability and accuracy of each model for water level forecasting **[113]. Shrike et al. (2012)** ANN and Fuzzy logic System based Model was generated. ANN based result applied in fuzzy system for further analysis. Fuzzy system was applied for opening and closing of the Dam gate. This Model is known as Neuro-Fuzzy model is to be prepared for Dam –Gate operation for flood monitoring and control system **[114]**

8. MODEL EVALUATION
On the basis of literature survey, it is observed that the evaluation of each approaches for both testing and training data are coefficient of determination R^2, root mean square error (RMSE), relative mean error (RME) and relative absolute error (RAE). These are widely implemented for the evaluation of effective results of time series forecasting [4-5].

9. CONCLUSION AND FUTURE STUDY
Based on the literature review, it is find out that data based models are an alternative method used for runoff forecasting particularly useful in flood prone area. Multiple linear regression (MLR) and autoregressive moving average (ARMA) models, autoregressive integrated moving average (ARIMA) are probably the most common data based methods currently employed for forecasting runoff . Data based artificial intelligence models like ANN; SVM can also be developed with limited data and without a priori knowledge of the mathematical relationships interlinking inputs with outputs. Based on the review, it is observed that available models have limitations of times; effective only for site specific, expensive in data collection, lack of security and data transferability in similar watershed. After critical review of the literature, it is observed that there are very few mathematical and computational models available to predict real time hydrological forecast objectives i.e., to provide livelihood security through flood management especially designed for degraded and flood prone areas. Also there is scarcity of model for prediction of livelihood security with respect to holistic development of environmental economic and socio-economic yield.

There are several potential directions for future research. First we would like to investigate what measures best reflect the performance of existing models. Second we try to develop a rainfall-runoff model and improve the hydrological prediction with maximum accuracy.

10. REFERENCES
[1] Chaubey V., Mishra S. and Pandey,S. K. (2014), Time series data mining in real time surface runoff forecasting through support vector machine, IJCA(0975-8887), ISBN : 973-93-80882-93-2, 98(3), PP 23-28. Doi> 10.5120/17163-7223. ISBN: 973-93-80882-93-2.

[2] Gupta, P., Mishra, S., and Pandey, S. K. (2014), Time series data mining in rainfall forecasting using artificial neural network, IJSET (ISSN: 2277-1581), 3(8), pp 1060-1065.

[3] Mishra,S., Tiwari,H. L., Shukla J. P. and Purvia,R. (2014), Estimation of runoff and flood risk in the Narmada River Basin using hydrological time series data mining, in 19th International conference on "Hydraulics, Water Resources, Coastal & Environmental Engineering (HYDRO 2014 INTERNATIONAL)", MANIT, Bhopal.

[4] Mishra, S. Gupta, P., Pandey, S.K., Shukla, J.P. (2014), An

Efficient Approach of Artificial Neural Network in Runoff Forecasting, IJCA, 92(5), PP 9-15. Doi> 10.5120/16003-4991.

[5] Mishra,S., Chaubey,V., Pandey,S. K . and Shukla,J. P. (2014), "An efficient approach of Support vector machine in runoff forecasting", IJSER (0975-8887), ISSN 2229-5518 , 5(3) PP. 158-167.

[6] Mishra, S., Dwivedi, V.K., Sarvanan, C. and Pathak, K. K. (2013), "Pattern Discovery in Hydrological Time Series Data Mining during the Monsoon Period of the High Flood Years in Brahmaputra River Basin", IJCA(0975-8887), doi-10.5120/11397-6698, 67(6), pp 7-14.

[7] Mishra,S., Saravanan,C.and Dwivedi,V. K. (2014), Estimation of flood magnitude and flood risk in the Brahmaputra river basin using hydrological time series data mining in the Brahmaputra river basin using hydrological time series data mining, International conference *on decision support systems for early warning and mitigation of disaster* (DSS-EWMD), NIT Durgapur.

[8] Mishra,S., Sarvanan,C., Dwivedi,V. K. and Shukla,J. P. (2014), Rainfall prediction using hydrological time series data mining, published in national workshop on "Technologies for Sustainable Rural Development-Having Potential for Socio-economic Upliftment" during July, 04-05, 2014 at CSIR-AMPRI, Bhopal, pp. 403-408.

[9] Mishra,S., Shukla,J. P., Saravanan,C., Dwivedi V. K. and Pathak,K. K. (2013), "An Innovative Approach of Data Mining Techniques in Stream Flow Forecasting" ISCA-ISC-2013-5CITS-05, Souvenir of 3rd International Science Congress, Coimbatore, India, 8th – 9th .

[10] Mishra,S., Majumder S. and Dwivedi V. K. (2011), "PATTERN DISCOVERY IN HYDROLOGICAL TIMESERIES DATA MINING " in a Sustainable Water resources Management And Climate Change Adaptation, II, pp.107-115, NIT Durgapur.

[11] Mishra, S., Sarvanan, C. ,Dwivedi, V.K., and Pathak, K. K. (2014), "Discovering Flood Recession Pattern in Hydrological Time Series Data Mining during the Post Monsoon Period", IJCA(0975-8887), ISBN : 973-93-80880-78-3, 90(08), PP 35-44 doi-10.5120/15597-4375.

[12] Mishra, S., Sarvanan,C., Dwivedi, V. K. and Pathak,K. K. (2015), "Discovering Flood Rising Pattern in Hydrological Time Series Data Mining during the Pre Monsoon Period", Indian. J.of Zeo-Marine Science, 44(3).

[13] Purviya,R., Tiwari,H. L. Mishra,S. (2014), "Application of Clustering Data Mining Techniques in Temporal Data Sets of Hydrology: A Review" , IJSET(ISSN:2277-1581) , 3(4), pp 360-365.

[14] Mohammadi, K., Eslami, R.H. and Kahawita, R. (2006), Parameter estimation of an ARMA model for river flow forecasting using Goal programming, ELSEVIER Sciences, Journal of Hydrology 331, pp 293-299.

[15] Abhishek, K., Kumar, A., Ranjan, R., and Kumar, S. (2012), A Rainfall Prediction Model using Artificial Neural Network, IEEE.

[16] Abrahart, R. J., See, L. & Kneale, P. E. (1999), Using pruning algorithms and genetic algorithms to optimise network architectures and forecasting inputs in a neural network rainfall-runoff model, journal of hydroinformatics.

[17] Adamowski, J. and Karapataki, C. (2010), Comparison of Multivariate Regression and Artificial Neural Networks for Peak Urban Water-Demand Forecasting: Evaluation of Different ANN Learning Algorithms, ASCE

[18] Adamowski, J. F. (2008), Peak Daily Water Demand Forecast Modeling Using Artificial Neural Networks, ASCE.

[19] Mittal, P., Chowdhury, S., Roy, S., Bhatia, N., and Srivastav, R. (2012), Dual Artificial Neural Network for Rainfall-Runoff Forecasting, Journal of Water Resource and Protection, 4, pp 1024-1028.

[20] Abbasi,M. Abduli, M.A., Omidvar, B. and Baghvand, A. (2013), Forecasting Municipal Solid waste Generation by Hybrid Support Vector Machine and Partial Least Square Model, Int. J. Environ. Res., 7(1), pp 27-38, ISSN: 1735-6865.

[21] Ramoni, M.,Sebastiani, P. and Cohen, P. (2000), Multivariate clustering by dynamics, Proceedings of the 2000 National Conference on Artificial Intelligence (AAAI-2000), San Francisco, CA, pp 633–638.

[22] Wijk, J. J. van and Selow, E. R. Van (1999), Cluster and calendar based visualization of time series data, Proceedings of IEEE Symposium on Information Visualization, San Francisco, CA, pp25–26.

[23] Vlachos, M.,Lin, J. and Keogh, E. (2003), A wavelet based anytime algorithm for k-means clustering of time series, Proceedings of the Third SIAM International Conference on Data Mining, San Francisco, CA, May 1–3.

[24] Li, C. And Biswas, G. (1999), Temporal pattern generation using hidden Markov model based unsupervised classification, in: D.J. Hand, J.N. Kok, M.R. Berthold (Eds.), Lecture Notes in Computer Science, vol. 164, IDA '99, Springer, Berlin, pp 245–256.

[25] Liao, T. W. (2005), Clustering of time series data—a survey, Pattern Recognition Society.

[26] Agarwal, P., Alam Afshar, M. And Biswas, R. (2010), Analysing the agglomerative hierarchical Clustering Algorithm for Caegorical Attributes, IJJMT, 1(2), pp 186-190.

[27] Hall, M.J., Minns, A.W. and Ashrafuzzaman, A.K.M., (2002), The application of data mining techniques for the regionalization of hydrological variables, Hydrology and Earth System Sciences, 6(4), 685–694.

[28] Mujumdar, P. P. and Kumar, D. N., 1990 "Stochastic models of streamflow: some case studies", *Hydrological Sciences Journal*, vol. 35,4.

[29] Wang, C. W., Chau, W. K., Cheng, T. C. and Qiu, Lin (2009), A comparison of performance of several artificial intelligent methods for forecasting monthly discharge time series, Journal of Hydro, 374(3-4), pp 294-306.

[30] Makridakis, S. and Hison, M., 1995 "ARMA MODELS and The Box Jenkins methodology" by revised version of 95/33/TM.

[31] Agarwal, A. and Singh, R.D. (2004), Runoff Modelling Through Back Propagation Artificial Neural Network With Variable Rainfall-Runoff Data, Water Resources Management, 18, pp 285–300.

[32] Boudaghpour, S., Bagheri, M., and Bagheri, Z.,2014,

"Using Stochastic Modeling Techniques to Predict the Changes of Total Suspended Solids and Sediments in Lighvan Chai Catchment Area in Iran", *Journal of River Engineering*, Vol. 2 , Issue 1.

[33] Tesfaye, Y. G., Meerschaert, M. M. and Andersonet, P. L., 2006 "Identification of periodic autoregressive moving average models and their application to the modelling of river flows", *Water Resources Research*, Vol. 42, p.p. no 1-11.

[34] Musa, J. J, 2013 "Stochastic Modelling of Shiroro River Stream flow Process", *AJER*, Vol-02, Issue-06, pp-49-54.

[35] E. Toth, Brath, A., and Montanari, A., 2000 "Comparison of short-term rainfall prediction models for real-time flood forecasting", *Journal of Hydrology*, Vol -239, p.p. no.132–147.

[36] Chakraborty, S., Denis, D.M., and Sherring, A. "Development of Time Series Autoregressive Model for prediction of rainfall and runoff in Kelo Watershed Chhattisgarh", *International Journal of Advances in Engineering Science and Technology*, Vol- 2, p.p. no. 153-163.

[37] Devi, C. J., Reddy, B. S. P., Kumar, K. V., Reddy, B. M., and Nayak, N. R. (2012), ANN Approach for Weather Prediction using Back Propagation, IJETT.

[38] Dreiseitl, S., and Ohno-Machado, L. (2002), Logistic regression and artificial neural network classification models: a methodology review, SCIENCE DIRECT.

[39] Edossa, D. C., Mukand, S., Babel, Forecasting Hydrological Droughts Using Artificial Neural Network Modeling Technique.

[40] Hsieh, B. B., Bartos, C. C. L., and Zhang, B. , USE OF ARTIFICIAL NEURAL NETWORKS IN A STREAMFLOW PREDICTION SYSTEM.

[41] Kumar, R., and Yadav, G. S. (2013), Forecasting of Rain Fall in Varanasi District, Uttar Pradesh Using Artificial Neural Network, JECET.

[42] Kumarasiri, A. D., and Sonnadara, U. J. (2008), Performance of an artificial neural network on forecasting the daily occurrence and annual depth of rainfall at a tropical site, Hydrological process, 22, pp 3535–3542

[43] Luc, K. C., Ball, J. E., and Sharma, A. (2001), An Application of Artificial Neural Networks for Rainfall Forecasting, ELSEVIER, 33(6–7), pp *683–693*.

[44] Machado, F., Mine, M., Kaviski, E., and Fill, H. (2011), Monthly rainfall–runoff modelling using artificial neural networks, Hydrological Sciences Journal, 56(3).

[45] Maier, H. R., Jain, A.,Dandy, G.C.and Sudheer, K. P. (2010), Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions, environmental modeling & software, 25, pp 891-909.

[46] Minns, A. W., and Hall, M. J. (1996), artificial neural networks as rainfall runoff models, Hydrological Sciences Journal.

[47] Mutlu, E., Chaubey, I., Hexmoor, H., and Bajwa, S. G. (2008), Comparison of artificial neural network models for hydrologic predictions at multiple gauging stations in an agricultural watershed, Hydological Process.

[48] Raman, H. and Sunilkumar, N. (1995), Multivariate modelling of water resources time series using artificial neural networks, Hydrological Sciences Journal, 40(2).

[49] Riad, S. Mania, J., Bouchaou, L., and Najjar, Y. (2004), Rainfall-Runoff Model Using an Artificial Neural Network Approach, Mathematical and Computer Modelling 40 pp 839-846.

[50] Sarkar, A. And Kumar, R. (2012), Artificial Neural Network for Event Based Rainfall-Runoff modelling, Journal of Water Resource and Prediction, 4, pp 891-897.

[51] Shamseldin, A. Y., Nasr, A. E. and O'Connor, K. M. (2002), Comparison of different forms of the Multi-layer Feed-ForwardNeural Network method used for river flow forecasting, Hydrology and Earth System Sciences, 6(4), pp 671–684.

[52] Sharda, V. N., Patel, R. M., Prasher, S.O., Ojasvi, P.R., and Prakash, C. (2006), Modeling runoff from middle Himalayan watersheds employing artificial intelligence techniques, agricultural water management 83, pp 233 – 242.

[53] Srinivasulu, S. and Jain, A. (2006), A comparative analysis of training methods for artificial neural network rainfall-runoff models, applied soft computing, 6, pp 295-306.

[54] Wu, J.S., P.E., Asce, M., Han, J., Annambhotla, S. and Bryant, S. (2005), Artificial neural network for forecasting watershed runoff and stream flows, Journal of Hydrologic Engineering, pp 216-222.

[55] Cigizoglu, H. K. (2005), Generalized regression neural network in monthly flow forecasting, Civil Engineering and Environmental Systems, 22 (2).

[56] Cigizoglu, H. K., and Alp, M. (2006), Generalized regression neural network in modelling river sediment yield, ELSEVIER

[57] Koutsoyiannis, D. (2007), Discussion of "Generalized regression neural networks for evapotranspiration modelling, IAHS.

[58] Kisi, O. (2006), Generalized regression neural networks for evapotranspiration modelling, IAHS.

[59] Wang, Z. L., and Sheng, H. H. (2010), Rainfall Prediction Using Generalized Regression Neural Network: Case study Zhengzhou, IEEE.

[60] Cannas, B., Fanni, A., Sias, G., Tronci, S., and Zedda, M. K. (2005), River flow foreasting using neural networks and wavelet analysis, European Geosciences Union.

[61] Adamowski, J. and Sun, K. (2010), Development of a coupled wavelet transform and neural network method ofr flow forecasting of non-perennial rivers in semi-arid watershed, ELSEVIER

[62] Adamowski, J., Chan, H. F., Prasher, S. O., and Sharda, V. N. (2012), Comparison of multivariate adaptive regression splines with coupled wavelet transform artificial neural networks for runoff forecasting in Himalayan micro-watersheds with limited data, IWA.

[63] Adamouski, J., and Chan, H. F. (2011), A wavelet neural network conjunction model for groundwater level forecasting, ELSEVIER

[64] Belayneh, A., and Adamowski, J. (2012), Standard precipitation index drought forecasting using Neural

Networks, Wavelet Neural Networks, and Support Vector Regression, Applied Computational Intelligence and Soft Computing Volume 2012 Article ID 794061, 13 pages.

[65] Campisi-Pinto, S., Adamouski, J., and Oron, G. (2012), Forecasting UrbanWater Demand ViaWavelet-Denoising and Neural Network Models. Case Study: City of Syracuse, Italy, Springer, 26(12) pp 3539-3558.

[66] Partal, T., and Cigizoglu, H. K. (2009), Prediction of daily precipitation using wavelet–neural networks, IAHS.

[67] Ramana, R. V., Krishna, B., Kumar, S. R., and Pandey, N. G. (2013), Monthly Rainfall prediction using wavelet neural network analysis, Springer.

[68] Santos, C.A.G. and da Silva, G.B.L.(2014), Daily streamflow forecasting using a wavelet transform and artificial neural network hybrid models,HydrologicalSciences.JournalJournaldesSciencesHydrologiques,59(2),PP.312324,http://dx.doi.org/10.1080/02626667.2013.800944

[69] Wei, S., Yang, H., Song, J., Abbaspour, K., and Xu, Z. (2013), A wavelet-neural network hybrid modeling approach for estimating and predicting river monthly flows, IAHS.

[70] Aggarwal, N. & Aggarwal, K. (2012), A Mid-Point based K-mean Clustering Algorithm for Data mining, IJCSE, 4(06) .

[71] Botsis, D., Latinopulos, P. and Diamantaras, K. (2011), Rainfall-Runoff Modeling Using Support Vector Regression and Artificial Neural Networks, CEST2011- Rhodes, Ref No. XXX, Greece.

[72] Bhagwat, P. and Maity, R. (2013), Hydroclimatic stream flow prediction using Least Square Support Vector Regression, ISH Journal of Hydraulic Engineering, 19(3) , pp 320-328.

[73] Cimen, M. (2008), Estimation of daily suspended sediments using support vector machines, Journal of Hydrological Sciences, l.53 (3).

[74] Espinoza, M., Johan, A. K. Suykens and Bart De Moor. (2006), Load Forecasting using Fixed-Size Least Squares Support Vector Machines.

[75] Flint, E.L. and Flint, L. A. (2012), Downscaling future climate scenarios to fine scales for hydrologic and ecological modeling and analysis, Ecological Processes, A Springer Open Journal. http://www. ecologicalprocesses.com/content/1/1/2.

[76] Lai, H.C. and Tseng, H.M. (2010), Comparison of regression models, grey models, and supervised learning models for forecasting flood stage caused by typhoon events, Journal of the Chinese Institute of Engineers, 33(4), pp 629-634.

[77] Li, H.P., Kwon, H.H., Sun, L., Lall, Upmanu and Kao, J. J. (2010), A modified support vector machine based prediction model on streamflow at the Shihmen Reservoir, Taiwan, International Journal Of Climatology 30, pp 1256-1268.

[78] Liong, Y.S. and Chandrasekaran, S. (2002), Flood Stage Forecasting With Support Vector Machines, Journal of the American Water Resources Association, 38(1).

[79] Lin, Y.J., Cheng, T.C. and Chau, W.K. (2006), Using support vector machines for long term discharge prediction, Hydrological Science Journal, 51(4), pp599-612.

[80] Ma, Xixia, Ping, J., Yang, L., Yan, M. and Mu, H. (2011), Combined model of chaos theory, wavelet and support vector machine for forecasting runoff series and its application, IEEE 978-1-61284-340-7/11.

[81] Maity, R. and Khasid, S.S. (2009), Hydroclimatological approach for monthly streamflow prediction using genetic programming, ISH Journal of Hydraulic Engineering, 15(2).

[82] Lingras, P. and Butz, C. (2004), Interval set Classifiers using Support Vector Machines, IEEE, 0-7803-8376-1/04.

[83] Mamat, M. and Samad, A.S., Performance of Radial Basis Function and Support Vector Machine in Time Series Forecasting.

[84] Misra, D., Oommen, T., Agarwal, A., Mishra, S.K. and Thompson, M.A. (2009), Application and analysis of support vector machine based simulation for runoff and sediment yield, ELSEVIER Science: 103(2009), pp 527-535.

[85] Mountrakis, G., Im, J. and Ogole, C. (2010), Support vector machines in remote sensing: A review, ISPRS Journal of Photogrammetry and Remote Sensing, doi:10.1016/j.isprsjprs.2010.11.001.

[86] Solomatine, P.D. (2002), Computational intelligence techniques in modeling water systems: some applications, IEEE, 0-7803-7278-6/02.

[87] Shahbazi, N.A. and Pilpayeh, A. R. (2012), River flow forecasting using support vector machines, 14th ISCCBE, Moscow, Russia.

[88] Torabi, M. and Hashemi, S. (2012), A Data Mining paradigm to Forecast Weather Sensitive Short-Term Energy Consumption, IEEE 978-1-4673-1479-4/12.

[89] Wu, H.C., Ho, M.J. and Lee, T.D. (2004), Travel-Time Prediction with Support Vector Regression, IEEE Transactions on Intelligent Transportation Systems, 5(4), 1524-9050/04.

[90] Yang, H., Huang, K., King, I. and Lyu, R.M. (2009), Localized support vector regression for time series prediction, ELSEVIER Sciences Neurocomputing 72, pp 2659-2669.

[91] Yi, D., Wei, C. and Shengfeng, L. (2011), A new Regression Method Based on SVM classification, Eighth International Conference on Fuzzy Systems and Knowledge Discovery, 978-1-61284-181-6/11.

[92] Xian, M.G. and Zeng, Q.B., (2008), A novel evaluation method basing on support vector machines, IEEE 978-0-7695-3134-2/08.

[93] Yoon, H., Jun, C.S., Hyun, Y., Bae, O.G. and Lee, K.K. (2011), A comparative study of artificial neural network and support vector achiness for predicting groundwater levels in a costal aquifer, ELSEVIER Sciences, Journal of Hydrology No. 396, pp 128-138.

[94] Hwang, H. S, Ham, H. D. and Kim, H. J. (2012), A new measure for assessing the efficiency of hydrological data-driven forecasting models, Hydrological Sciences Journal, 57(7), pp 1257-1274.

[95] Samsudin, R., Ismail, S. and Shabri, A. (2011), "A Hybrid Model of Self-Organizing Maps (SOM) and Least Square Support Vector Machine (LSSVM) for Time-Series Forecasting," Expert Systems with Applications, 38(8),

pp. 10574-10578.

[96] Shabri, A. and Suhartono, (2012), Streamflow forecasting using least-square support vector machines, Hydrological Sciences Journal, 57 (7), pp 1275-1293.

[97] Shah Shiloh, R. (2005), Least Squares Support Vector Machines

[98] Suykens, K.A.J, Brabanter, D.J., Lukas, L. and Vandewalle, J. (2001), Weighted least support vector machines: robustness and sparse approximation, ELSEVIER Science Neurocomputing: 48, 0925-2312/02, pp85-105. Doi: 10.1023/A: 1018628609742.

[99] Ismail, S., Shabri, A. and Samsudin, R. (2012), A hybrid model of self organizing maps and least square support vector machine for river flow forecasting, Hydrology and Earth System Sciences, 16, pp 4417-4433.

[100]Lin, F., Yeh, C.C. and Lee, Y.M. (2013), A Hybrid Business Failure Prediction Model Using Locally Linear Embedding and Support Vector Machines, Romanian Journal of Economic Forecasting.

[101]Pandhiani, M.S. and Shabri, B.A. (2013), Time Series Forecasting Using Wavelet-Least Squares Support Vector Machines and Wavelet Regression models for Monthly Stream Flow Data, Open Journal of Statistics,3, pp183-194. doi: 10.4236/ojs.2013.33021.

[102]Sivakumar, B., Berndtsson, R., Olsson, J., Jinno, K. and kawamura, A. (2000), Dynamics of monthly rainfall-runoff process at the Gota basin: A search of chaos, Hydrology and Earth System Sciences, 4(3), pp 407-417.

[103]Mutao, H. (2007), An Intelligent Hybrid Genetic Annealing Neural Network Algorithms for Runoff Forecasting, ASCE.

[104]Wang, W., Xu, D. and Qiu, L. (2010), Support Vector Machine with Chaotic Genetic Algorithms for Annual Runoff Forecasting, ICNC IEEE, 978-1-4244-5961-2/10.

[105]Wang, W. C., Xu, D. M., Chau, K. W. and Chen, S. (2013), Improved annual rainfall-runoff forecasting using PSO–SVM model based on EEMD, Journal of Hydroinformatics, doi:10.2166/hydro.2013.134.

[106]Wei, L.H. and Billings, A.S. (2006), Long term prediction of non-linear time series using multiresolution wavelet models, International Journal of Control, 79(6), pp 569-580.

[107]Yu, X., Liong, Y.S. and Babovic, V. (2004), EC-SVM approach for real time hydrological forecasting, Journal of Hydroinformatics, 06.3.

[108]Li, S., Qi, R. & Jia, W. (2009), Calibration Of The Conceptual Rainfall-Run Off Model's Parameters, Advances in Water Resources and Hydraulic Engineering, pp 55-59.

[109]Patil, S., Patil, S. & Valunjkar, S. (2012), Study of Different Rainfall-Runoff Forecasting Algorithms for Better Water Consumption, ICCTAI.

[110] Banik, S., Anwer, M., Khan, A. F. M. K., Rouf, R. A. & Chanchary, F. A. (2009), Forecasting Bangladeshi monsoon rainfall using neural network and genetic algorithm approaches, ITMR.

[111] Dong, S., Zhou, H., and Xu, H. (2004), A Forecast Model of Hydrologic Single Element Medium and Long-Period Based on Rough Set Theory, Water Resources Management, 18, pp 483-495.

[112]Mahabir, C., Hicks, F. And Fayek, A. (2003), Application of fuzzy logic to forecast seasonal runoff, Hydrological Processes, 17, pp 3749-3762.

[113] Alvisi, S., Mascellani, G., Franchini, M. and Bardossy, A. (2006), Water level forecasting through fuzzy logic and artificial neural network approaches, Hydrology and Earth System Sciences, 10, pp 1-17.

[114]Shirke, Y., Kawitkar, R. And Balan, S. (2012), Artificial Neural Network based Runoff Prediction Model for a Reservoir, IJERT, 1(3), pp 1-4.