Pattern Discovery in Hydrological Time Series Data Mining during the Monsoon Period of the High Flood Years in Brahmaputra River Basin

Satanand Mishra,	V. K. Dwivedi	C. Saravanan	K. K. Pathak
Scientist	Professor	Asstt. Professor	Professor
IT& WRM Group, CSIR-AMPRI	Dept of Civil Engineering,	Computer Centre	Dept of C&E Engg.,
Bhopal- 462064, India	NIT,Durgapur-713209, India.	NIT, Durgapur-713209, India	NITTTR Bhopal, 462002, India

ABSTRACT
In this paper, based on data mining techniques, the analysis is carried out in hydrological daily discharge time series of the Panchratna station in the river Brahmaputra under Brahmaputra and Barak Basin Organization in India. The data has been selected for the high flood years 1988, 1991,1998, 2004, and 2007. The whole year is divided into three periods known as Pre-monsoon, Monsoon and Post Monsoon. In this paper, only monsoon period data have been used. For standardization of data, statistical analysis such as mean monthly discharge, monthly Maximum Discharge, monthly amplitude and monthly standard deviation have been carried out. K-means clustering is segmented for the monsoon period process of daily discharge. Dynamic Time Warping (DTW) is used to look for similarities in the discharge process under the same climatic condition. Similarity matrix helped in the mining of discharge process in similar time period in the different years. The agglomerative hierarchical clustering is used to cluster and discover the discharge patterns in terms of the autoregressive model. A forecast model has been predicted on the discharge process.

Keywords : data mining; hydrological time series; clustering; pattern discovery , similarity search.

1. INTRODUCTION
Worldwide water resources organizations are engaged in hydrological and hydro-meteorological management and observing water level, rainfall, discharge, sediment, evaporation, temperature and water quality data. These data are very useful in research, historical trend analysis and future forecasting. With the development of database technology, various techniques of data analysis and knowledge extraction tools are being used for collected time series data in scientific as well as commercial organizations. Data mining, also referred to as Knowledge Discovery in Database (KDD) is defined as "Discovery of comprehensible, important and previously unknown rules or anything that is useful and non-trivial or unexpected from our collected data [15]". Today, data mining is being vastly applied in research and business field. Finding association rules, sequential patterns, classification and clustering of data are typically task involved in the process of data mining. Mainly, data mining is an iterative process in which data have to be critically selected and cleaned; parameters of the mining algorithms are familiarised [10]. The nature and quality of collected data in hydrology are extremely important and all characteristics of such data should be the best possibly analysed. Data mining in Hydrology depends on the hydro-meteorological data, which generally takes the form of time series. Hydrological time series are sets of various record values of hydrological data that vary with time [11].

In the field of hydrological forecasting, with the help of data mining techniques, earlier various researches have been carried out. Some of them are : Similarity search and pattern discovery in hydrological time series data mining [14], Flood pattern detection using sliding window technique [17], Applications of Data Mining in Hydrology [11], Runoff forecasting using fuzzy support vector regression [23], Forecasting monthly runoff using wavelet neural network model [2], Neural network model for hydrological forecasting based on multivariate phase space reconstruction [22], Mid-short term daily runoff forecasting by ANNs and multiple process based hydrological models [7], Research and application of data mining for runoff forecasting [10], River flow time series using least squares support vector machines[18], a novel approach to the similarity analysis of multivariate time series and its application in hydrological data mining [25], Computational methods for temporal pattern discovery in biomedical genomic databases [16], an efficient k-Means clustering algorithm: analysis and implementation[9], The prediction algorithm based on fuzzy logic using time series data mining methods [3] , A forecast Model of Hydrologic Single, Element Medium and Long-Period Based on Rough Set Theory [19], Application of ANN in Forecast of surface runoff [4] are important contributions in the knowledge discovery from hydrological databases using time series data mining.

In this paper, Time Series Data Mining has been used for hydrological study [14]. Time series data mining combines the fields of time series and data mining techniques [3]. Here, the object is to develop a data mining application using modern information technology to discover the hidden information or patterns behind the historical hydrological data of the river Brahmaputra under the hydrological process, and also to look for a new way to meet the requirements of hydrological time series analysis using TSDM algorithms.

2. TSDM ALGORITHMS
Time series data are the classes of data where the sequence of values change during a period with time, for example the amount of sales, temperature changes, earthquake eruption, a patient's heart rate changes, financial stock sales, and hydrological observations as river water level, river discharge in particular stations and so on [14, 25]. Time series data mining algorithm can be used to predict continuous values of data. Once the algorithm is skilled to predict a series of data, it can predict the outcome of the other series of data. The algorithm generates a model that can predict trends based on only on the original data set. New data can also be added that becomes part of the trend analysis. Some mature TSDM algorithms comprise indexing, classification, clustering and segmentation of time series are applied in similarity search,

sequential pattern discovery and association rule mining, thus creating many new algorithms for TSDM. In our study, TSDM algorithms usage was restricted to: cluster analysis, similarity search and pattern discovery. There are two main goals of time series analysis: (i) identifying the nature of the phenomenon represented by the sequence of observations, and (ii) forecasting (predicting future values of the time series variable). Both of these goals require that the pattern of observed time series data is identified and more or less formally described. Once the pattern is established, we can interpret and integrate it with other data. Regardless of the depth of our understanding and the validity of our interpretation of the phenomenon, we can extrapolate the identified pattern to predict future events.

2.1 Cluster analysis

Clustering is one of the unsupervised learning methods. Its main goal is to identify structure in an unlabeled data set by objectively organizing data into homogeneous groups where the within- group-object similarity is minimized and the between-group-object dissimilarity is maximized [21]. The clustering is defined as process of organizing objects into groups whose members are similar in some way. In other way, 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 are very dissimilar to objects in other clusters. In recent years clustering of time series has received considerable attention because of its fundamental task in data mining. Clustering methods used in time series are K-means clustering, K-medoids clustering, nearest neighbour clustering, hierarchical clustering, self-organizing maps, and so on.

Among all clustering algorithms, K-means clustering is more useful for finding spherical-based clusters capability in small-to medium-sized databases[6]. K-means clustering algorithm applied in following steps. First, it selects k of the objects, each of which initially represents a cluster mean or centre. For each of the objects in the dataset that remain, an object is assigned to the cluster to which it is the most similar, based on the distance between the object and the cluster mean(Euclidean Distance or any other distance measures). As soon as the objects are assigned to the best fit cluster, the cluster means are updated. This process iterates until the cluster centres no longer make a movement.

Beside it another classical clustering algorithm is the hierarchical clustering method [1]. Hierarchical Clustering refers to the formation of a recursive clustering of the data points: a partition into two clusters, each of which is itself hierarchically clustered. It forms a cluster tree by grouping the data objects. Hierarchical methods can be classified as being either agglomerative or divisive, on the basis of how the hierarchical decomposition is been formed. The agglomerative approach or bottom–up approach starts with the forming a separate group of each object. It successively merges the objects or groups close to one another, until all of the groups merge into one, or until a termination condition holds. Whereas in case of the divisive approach or top–down approach, starts with all the objects in the same cluster. Then a cluster is split up into smaller clusters, until eventually each object is in one cluster, or until a termination condition holds. In the Fig 1 AHC is illustrated.

Steps in Agglomerative Hierarchical Clustering -

1. Start with N clusters each containing a single entity, and an N × N symmetric matrix of distances (or similarities) Let dij = distance between item i and item j.
2. Search the distance matrix for the nearest pair clusters (i.e., the two clusters that are separated by the smallest distance). Denote the distance between these most similar clusters U and V by d_{UV}
3. Merge clusters U and V into new clusters labelled T. Update the entries in the distance matrix by (a). Deleting the rows and columns corresponding to clusters U and V, and (b). Adding a row and column giving the distances between the new cluster T and all the remaining clusters.
4. Repeat steps (2) and (3) a total of N-1times.

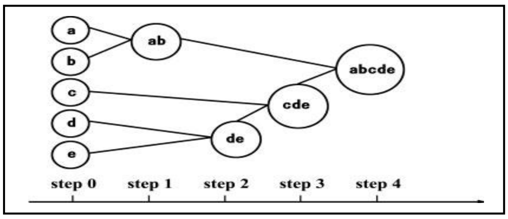


Fig 1: Cluster tree obtained by AHC

2.2 Similarity search

The similarity search finds data sequences in time series that differ only slightly from the given query sequence. It can be classified into two categories- i) Whole matching: In this kind of matching the time series data has to be of equal length and ii) Subsequence matching: In this mentioned category of matching, a query sequence X and a longer sequence Y. The objective is to identify the subsequence in Y, beginning at Yi, which best matches X, and report its offset within Y [13].

For similarity analysis of time series data, Euclidian distance is typically used as a similarity measure. Given two sequences are as follows :

X = $(x_1, \ldots, x_n)$ and Y = $(y_1, \ldots, y_m)$ with n = m, their Euclidean distance is defined as follows:

$$D(X,Y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \qquad (1)$$

2.3 Dynamic time warping (DTW) algorithm:

DTW is an algorithm for measuring optimal similarity between two time data sequences [8-5]. The time series data varies not only on the time amplitudes but also in terms of time progression as the hydrological processes may reveal with different rates in response to the different environmental conditions. A non-linear alignment produces a similar measure, allowing similar shapes to match even if they are out of phase in time axis. The sequences are "warped" non-linearly in the time dimension to determine a measure of their similarity independent of certain non-linear variations in the time dimension. To find the best alignment between time sequences X&Y one needs to find the path through the grid.

2.4 Pattern discovery

Discovery of patterns in data mining is a lucrative and highly demanding work. Data are sampled over time as $X=X_1,X_2,X_3\ldots X_t,..X_l$ (where l=length of data and the t denotes the sample). X are not independently and identically distributed. The X may come from different processes dependent on each other. Pattern discovery aims to find a subset of data from the available dataset, such that the subset represents a trend in the data. This trend when detected and modeled by an equation can be used in forecasting future responses of data. The problems that arise with detecting patterns are that, the data may contain multiple patterns, the data might be multidimensional, Even automated pattern discovery is difficult when the time series data is lengthy. In this case, discovered time series patterns, predict the future behavior of data that changes with time. This is a scope of trend analysis and prediction in time series data analysis [12].

2.5 Auto regressive model (ARM)

An Auto Regressive Model is a forecast model. The autoregressive model is one of a group of linear prediction formulas that attempt to predict an output y[n] of a system based on the previous outputs

(y[n-1],y[n-2]...) and inputs (x[n], x[n-1], x[n-2]...)

Deriving the linear prediction model involves determining the coefficients a1, a2, and b0, b1, b2, in the equation:

y[n] (estimated) = a1.y[n-1] + a2.y[n-2]... + b0.x[n] + b1.x[n-1] + ...) (2)

The remarkable similarity between the prediction formula and the difference equation used to describe discrete linear time invariant systems. Calculating a set of coefficients that give a good prediction y[n] is tantamount to determining what the system is, within the constraints of the order chosen. A model which depends only on the previous outputs of the system is called an autoregressive model (AR), while a model which depends only on the inputs to the system is called a moving average model (MA), and of course a model based on both inputs and outputs is an autoregressive-moving-average model (ARMA). By definition, the AR model has only poles while the MA model has only zeros. Several methods and algorithms exist for calculating the coefficients of the AR model [24].

3. DATA MINING IN HYDROLOGICAL TIME SERIES

The process of data mining in hydrological time series involves following steps:

1. Calculation of four statistical Characteristics eg. Q_{mean},Q_{max},Q_{range}, and Q_{dev}

2. Standardization of these characteristics using Z-scores methods

3. Clustering Monthly Discharge process using k –means clustering algorithms

4. Segmentation of hydrologic periods of the annual process

5. Use dynamic time warping algorithm and detecting similarities in discharge process in Monsoon Period,

6. Applying agglomerative hierarchical clustering algorithms and discover pattern of discharge process

7. Analyzing the causal-effect relationship

8. Forecasting by Auto Regressive predictive model.

4. APPLICATION OF TIME SERIES DATA MINING

4.1 Selection of sites and data set used

The data in this work refer to that of the river Brahmaputra and the site was Panchratna (Latitude 26° 11’ 55" and Longitude 90° 34’ 38”), located in the district of Goalpara in the state of Assam shown in Figure 2. The length of the river upto the site is 2562 Km. The catchment area upto the site measures 468790 sq km. The site is located on the left bank of river. The type of site is HO (Hydrological Observation site). For this study the daily discharge, water level data for the entire year was taken during the highly flooded years of 1988, 1991,1998,2004,2007.

Fig 2: Google map: view of the basin and catchment

4.2 Data preparation

The four statistical characteristics (Q_{mean},Q_{max},Q_{range},Q_{dev}) were obtained for each month in discharge data as formula is given in table1. There were five years accounted viz. 1988,1998,2001,2004,2007.In order to have an effective analysis the data were standardized using Z-scores technique so that the mean of the entire data range leads to 0 and the standard deviation is 1. The need for standardization was that to avoid affecting the study results by the wide variations in the data. The reason for such preference is that calculating *z* requires the discharge mean and the discharge standard deviation, not the sample mean or sample deviation.

Table 1. Statistical characteristics of monthly discharge

Statistical Characteristics	Formula
Mean monthly discharge	Q_{mean} = $(1/n)\sum^{n}_{i=1} x_i$
Monthly maximum discharge	Q_{max} = $Max(x_1,x_2\ldots x_n)$
Monthly minimum discharge	Q_{min} = $Min(x_1,x_2\ldots,x_n)$
Monthly discharge range	Q_{range} = Q_{max}- Q_{mean}
Standard deviation discharge	Q_{std} = $(1/n)\sum^{n}_{i=1} (x_i - Q_{mean})^2$

4.3 Data segmentation

The statistical attributes obtained were subjected to K-means clustering. A total of 60 (12*5=60) cases based on 4 attributes from data of 5 years were used. The specifications taken during clustering were clustering criterion (Determinant), Initial Clusters (Random Generated), Repetitions 10. According to the cases distribution in the clusters, the annual discharge process could be obtained as separate 3 clusters segments: Period I (PreMonsoon: Jan-May), period-II (Monsoon: Jun-sep), period-III (Post Monsoon (Oct-Dec). Fig.-3 shows the 3 Hydrological Periods.

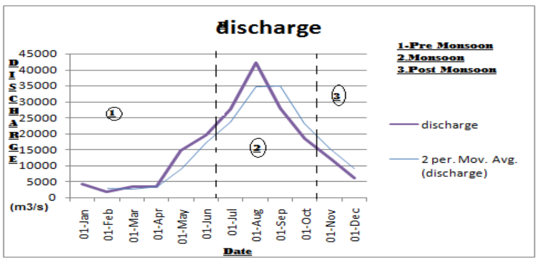


Fig 3: Three hydrological periods

It can be aptly inferred that the graph shows the highest discharge in the monsoon periods. This can be accounted due to the rainfall. In the pre-monsoon months which comprise the season of winter the graph suffers a low stature, from the month of April the curve starts rising, with a sharp angle denoting a transition in the month of May. August shows a peak in the discharge curve (the mid monsoons), accounting the heavy rainfall as reason for the high discharge, followed by a rapid decline and recession in the Post-Monsoon period of Oct-Dec. The discharge also gradually changes with a decrease in temperature in September. The different climates cause different discharge processes, thus for a better study the discharge processes must be studied under same formation mechanism. In his paper, we have focused only the Monsoon period.

4.4 Detecting similarity

The next course of work involved finding similar discharge processes. For this we had at our disposal 5 sets of data corresponding to each of the highly flood years (1988, 1998, 2001, 2004, and 2007). The DTW search technique was used. This is because time series are expected to vary not only in terms of expression amplitudes, but also in terms of time progression, since flow of water may unfold with different rates in response to different natural conditions or within different locations in the basin in different times. A matrix (M) is obtained of size 5X5, and the (i_{th}, j_{th}) element denotes the distance between the discharge processes for the i^{th} and j^{th} year.

A simulator was designed named the DTW Matrix Viewer in Java^{R} which gave the DTW similarity matrix for the data as given in figure 4. A user was allowed to select the corresponding hydrological period, which was obtained through previous work of clustering. The work of the simulator was to produce the DTW similarity matrix for the years with discharge data corresponding to the months in the hydrological period. The similarity matrix gave a comparison of the discharge processes in two years.

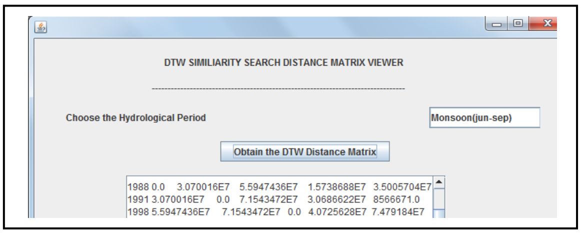


Fig-4: DTW Matrix viewer-showing the matrix

In the above simulator, inputs are discharge data for years 1988, 2007 ranging from month of June-Sep. Repeating the process for one year against all other years for a period gives the distance values for discharge time series data for two years. Proper iteration, representation generates the matrix as given in Table 2. The Matrices were generated for the three periods, amongst which only Monsoon is shown.

Table 2. DTW similarity matrix for the monsoon period (Jun-Sep)

	1988	1991	1998	2004	2007
1988	0.0	3.07	5.59	1.57	3.50
1991	3.07	0.0	7.15	3.06	3.3
1998	5.59	7.15	0.0	4.07	7.47
2004	1.57	3.06	4.07	0.0	5.53
2007	3.50	3.3	7.47	5.53	0.0

In this Matrix, the diagonal elements are 0 which can be aptly detected from the fact that a year is similar to the same year. The lowest value is of 1988 and 2004, which means in the hydrological periods they are the most similar years. Similarly, the matrices for other Hydrological Period can be obtained.

For the Monsoon Jun-Sep the discharge pattern was similar in the years 1988--2004, then 1991-2004, 1988-1991, 2007-1988, and 1998-2004.

On the basis of the above matrices the similarity graphs were obtained to have an idea about the similar discharge processes in the corresponding two years are given in the Fig 5 and Fig 6.

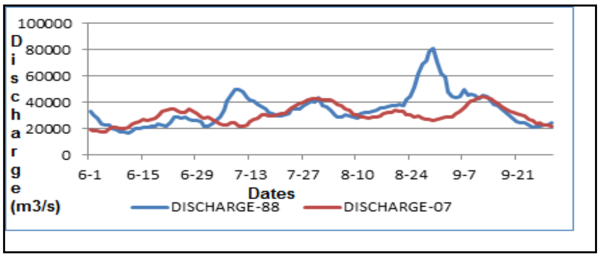


Fig 5: Similarity in discharge process of monsoon (Jun-Sep) 1988, 2007

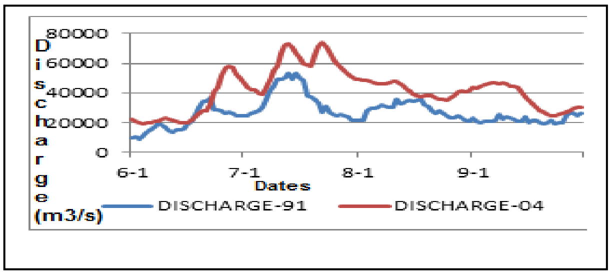


Fig 6: Similarity in discharge process of Monsoon (June-Sep) in 1991, 2004

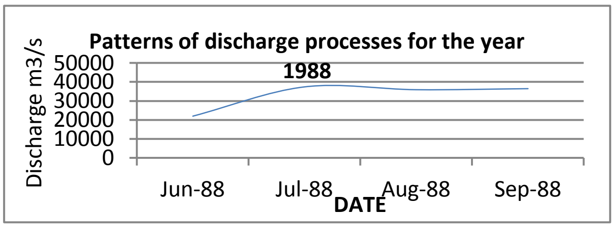
4.5 Discovering patterns

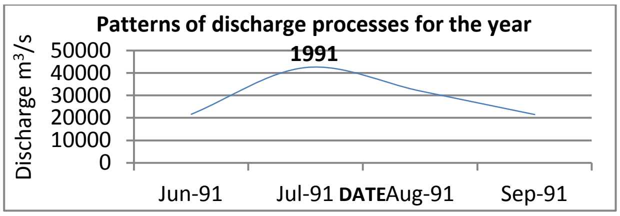
The next task was to identify the discharge pattern from the corresponding discharge time series data. For this, each hydrological period were obtained after the segmentation from the k- means clustering and then the discharge pattern in each of the periods was detected. The analysis is involved in the hierarchical clustering techniques on the 5 years as the cases/samples and the attributes as observations of the average discharge data for the months which are in the hydrological period has done. The analysis of the patterns in the data (June, July, Aug, Sep) as the attributes monsoon period were taken for the 5 years as the samples and the observations of discharge.

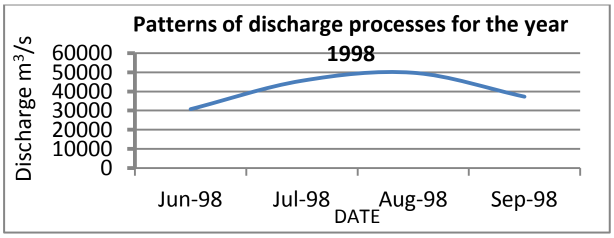
HC algorithm is particularly useful to find undetected patterns in the multidimensional data. As it is an unsupervised learning scheme, the number of clusters may be large or small at times. The lead role of HC is to find identify clusters or groups of related discharge time series natures that are similar to each other (objects). Now the discharge time series data of the cluster center is the pattern because all other objects in a particular group show similarity to the center only. Thus the cluster center can be taken as the pattern of discharge. On the basis of these five patterns, we have found the standard pattern for the discharge during the high floods. This is the similar to all received pattern for a particular year.

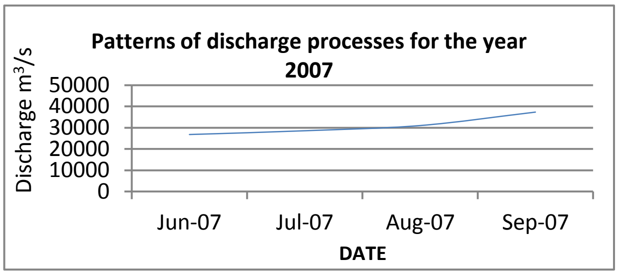
In this analysis, the discharge pattern (time series data of discharge for the months in the hydrological period) is cluster of a year (among 5 years) into several clusters. But the year which formed the cluster center, formed the pattern with its discharge data for the months in the period. All other members (years) in the cluster attained membership of the cluster because there was similarity to the year representing the center, so they can be said to follow the pattern.

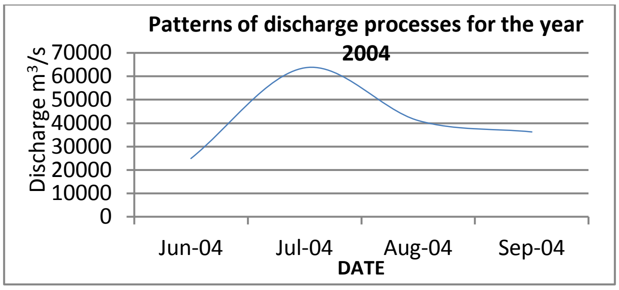
Now consider the center (the year) in the cluster and plotting of the discharge data of that year corresponding to daily discharge in the months, along the x-axis would give the pattern as shown Fig 7. In the Fig 7 the standard pattern was detected during the high floods at monsoon periods based on discharged data during the high floods year. The standard pattern shows the future patterns of the floods during the monsoon.











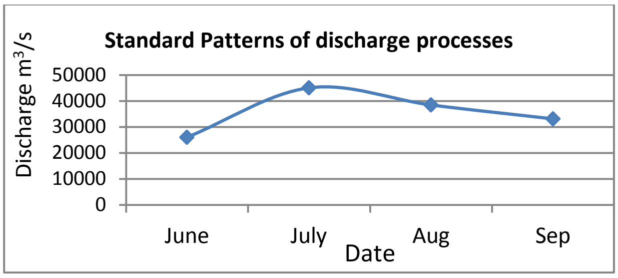


Fig 7: Patterns of discharge processes corresponding to all clusters obtained from AHC

One product of cluster analysis is a tree diagram representing the entire process of going from individual points to one big cluster. This diagram is called a dendrogram, and it is illustrated in Fig 8.

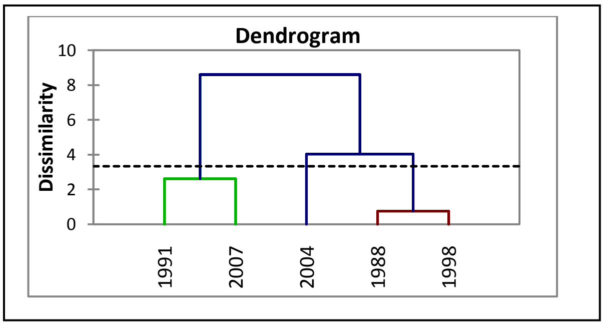


Fig 8: Dendrograms obtained after AHC of data for Monsoon, periods respectively.

In the dendrograms the height of each U shaped line denote the distance between the objects being connected with following parameters:

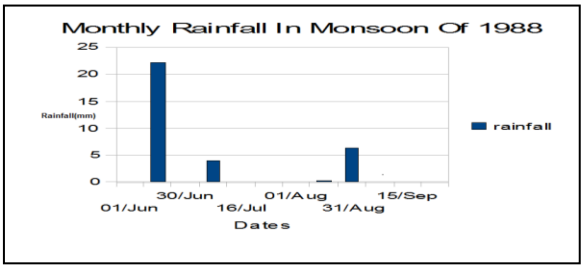
Specifications Taken During Hierarchical Clustering

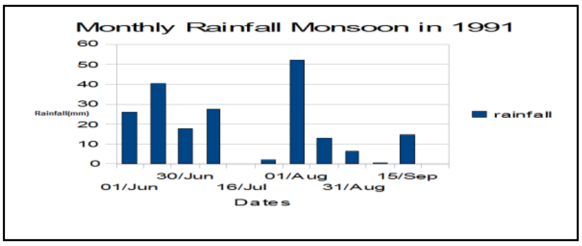
(i) Proximity type: Dissimilarities, (ii) Distance: Euclidean

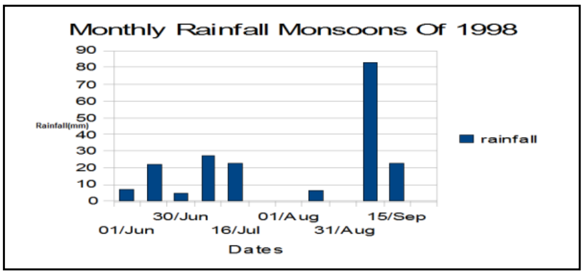
(iii) Agglomeration method: Ward's, and (IV) Cluster: along rows.

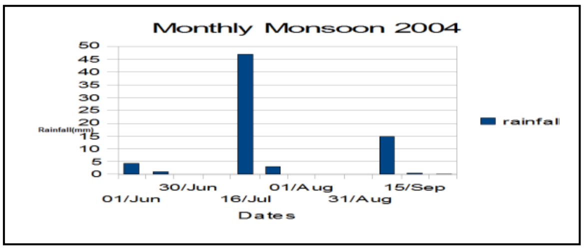
5. ANALYSIS AND INFERENCES

The hydrological processes show a causal-effect relationship with several happenings in nature as well as human interferences. As we can infer that the annual discharge processes is mainly affected by the rainfall, and the temperature too. The rainfall distribution was taken for 5 years during the monsoon period of Jun-September in the following Figures 9. The graphs shows similar pattern that conformed to the discharge pattern. This shows that rainfall is related strongly as contributor to discharge. The distribution of rainfall could be seen throughout for the period of monsoon in 1991.









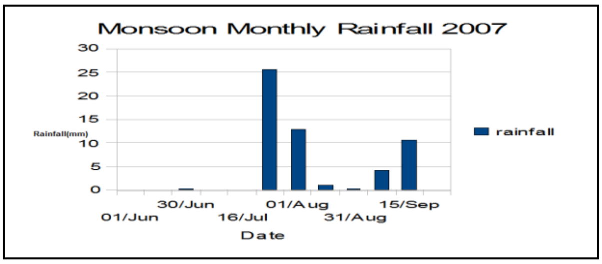
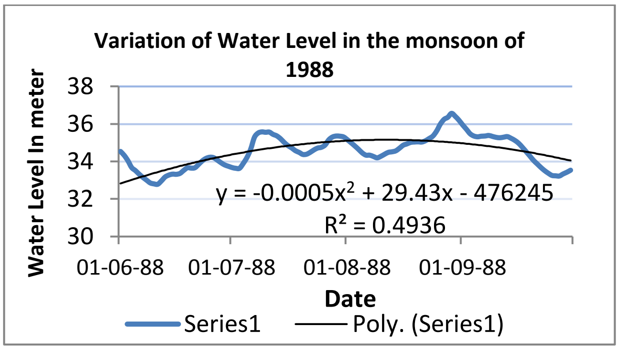
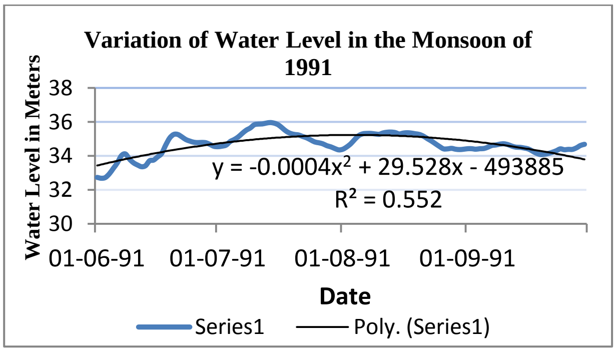
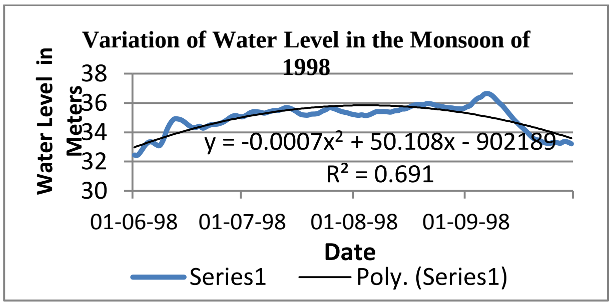


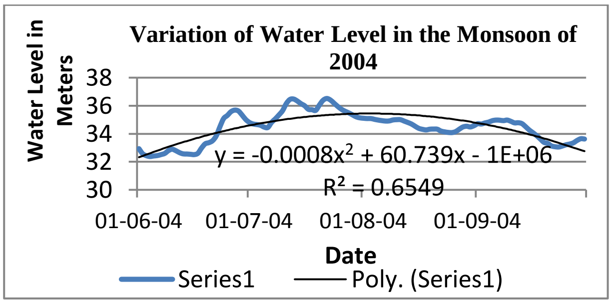
Fig 9: Monthly distribution of rainfall in the monsoons of the five highly floods years.

The relationship between the discharge and rainfall has been detected. For such analysis, the month of July selected for the year 1998 and found out the correlation factor (Pearson correlation) between the discharge and rainfall. The Pearson coefficient is independent of the scales and the units of variable measured. In the analysis the coefficient comes out to be 0.22 which is fairly good value to show the association of discharge with the rainfall.









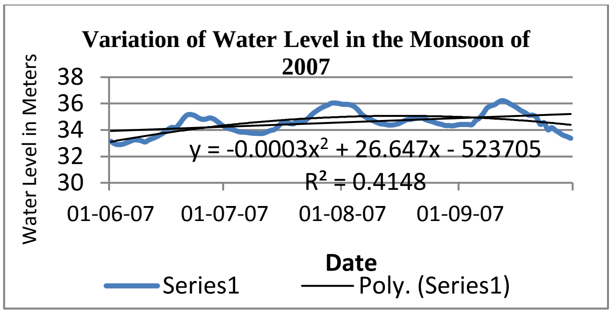


Fig 10: Hydrograph exhibiting the variation in water level in Monsoon periods of the highly flooded years

The variation of water level for the river Brahmaputra during the monsoon periods are shown in Figures 10. The above figure (Fig 7, Fig 9, and Fig 10) shows the very close relation in the variation of water level, rainfall and discharge pattern during the same periods. Water level, rainfall variation graphs, validates our discharge pattern. Polynomial series of water level is like to discharge pattern and R^2 values 0.414 to 0.691 shows the variation of water level up and down falls during the high floods in the monsoon period.

5.1 Inferences from the graph

The graph as expected shows an increasing nature from the month of June-Sep. From the end of May it started having a rise with a bend in the month of June. The horizontal line in deep blue gives the average water level. So in beginning of June, it crosses the average reading. With a sharp decline it crosses the average level in the end of June. In the beginning of July the curve is kind of flat indicating that it reaches a bit saturation, which is followed by sharp rises and declines the trend goes on. The dotted line gives the trend which is a linear ascending one.

6. FORECASTING MODELS

Based on the 5 patterns obtained, the predictive models has been developed using Auto Regressive modelling technique. AR(3) model of degree 3, based on lag predictor variables Q_{n-1},Q_{n-2}, Q_{n-3}. The y-intercept or the constant was nullified and not taken into account. The Table 3 enlists the models. The model is used for future prediction of discharged during the monsoon. The patterns extracted from hydrological data are valid for new hydrological data with some degree of certainty.

Pattern	Model
P1	Q_n=-0.06Q_{n-1}-0.6543Q_{n-2}+0.6072Q_{n-3}
P2	Q_n=-0.536544Q_{n-1}+0.71665Q_{n-2}-0.93774Q_{n-3}
P3	Q_n=0.9912Q_{n-1}+0.9912Q_{n-2}-0.9884Q_{n-3}
P4	Q_n=0.5526Q_{n-1}+0.4184Q_{n-2}+0.2951Q_{n-3}
P5	Q_n=0.899Q_{n-1}+0.819Q_{n-2}-0.747Q_{n-3}

Table 3. Forecasted models

7. CONCLUSIONS

In this paper, data mining techniques like clustering, similarity search and pattern discovery were used in hydrological discharge time series data and discovered the results as a similar discharge process and patterns. The comparison of hydrographs and rainfall during the same time period, it is proved that the discharge patterns were more similar under the same climatic periods. The patterns found by the AR Model can we used for the prediction of future value of discharge.

In Indian continent the whole river system is divided in three periods viz monsoon, post monsoon and pre-monsoon period. In this study, we have used only monsoon data during the high floods year. The future study will focus on post monsoon and pre monsoon data which will show the complete study of hydrological behaviour of river during high floods year for the particular stations.

8. ACKNOWLEDGEMENT

The authors would like to thank the Central Water Commission, Ministry of Water Resources for providing Water Level and Discharge data.

9. REFERENCES

[1] Anuradha, K., and Sairam, N. 2011, Classification of Images using JACCARD co-efficient and higher –order co-occurrences, JATTI, Vol.-34 No.1, 100-105.

[2] Aiyun, L., and Jiahai, L. 2011, Forecasting monthly runoff using weblet neural network model. International conference on Mechatronic Science, Electronic Engineering and Computer. IEEE. 978-1-61284-722-1 .

[3] Aydin, I., Karakose, and Akin, A., 2009, The prediction Algorithm based on Fuzzy logic using time series data mining method, World Academy of Science, Engg. and Technology, 51.

[4] Feng, L.H., Zhang, J.Z., 2010, Application of ANN in Forecast of surface runoff, IEEE (INC), 2010.

[5] Giorgino, T. 2009, Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package, Journal of Statistical Software, 31, Issue 7.

[6] Han, J.W., and Kamber, M. 2001, Data Mining Concept and techniques, Morgan Kaufman: San Fransciko, CA.

[7] Jingwen, XU. 2009, Mid-short term daily runoff forecasting by ANNs and multiple process based hydrological models, IEE, Conference : YC-ICT, 526-529.

[8] Kadir, A., and Peker, 2005, Subsequence Time Series (STS) Clustering Techniques for Meaningful Pattern Discovery, KIMAS 2005, April 18-21, Waltham, MA,USA, IEEE,0-7803-9013.

[9] Kanungo,T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R., and Wu, A.Y. 2002, an efficient k-Means clustering algorithm: analysis and implementation, IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 24, NO. 7, 881-892.

[10] Li, C., and Yuan, X. 2008, Research and Application of Data Mining for Runoff Forecasting, IEEE, 978-0-7695-3357-5/08.

[11] Liang, X., and Liang, Y. 2001, Applications of Data Mining in Hydrology, ICDM0, 1-0-7695-1119-8/01, 2001 IEEE.

[12] Liao, T. W. 2005, Clustering of time series data—a survey, Pattern Recognition, 38, 2005, 1857–1874.

[13] Ni, X. 2008, Research of Data mining based on neural Networks, WASET, 39, 2008.

[14] Ouyang, R., Ren, L., Cheng, W., and Zhou, C., 2010, "Similarity search and pattern discovery in Hydrological time series data mining, Wiley InterScience , Hydrol. Process,24, 1198-1210.

[15] Piatetsky-Shapiro, G., Frawley, W. J. 1991. Knowledge Discovery in Databases, AAAI/MIT Press: Boston, MA.

[16] Rafiq, M.I., Martin, J., Connor, O., and Das, A.K. 2005, Computational method for Temporal Pattern Discovery in Biomedical Genomic Database, IEEE(CSB'05), 0-7695-2344-7/05.

[17] Ruhana, K., Mohamud, K., Zakaria, N., Katuk, N., and Shbier M. 2009, Flood Pattern Detection Using Sliding Window Techniques, 978-0-7695-3648-4/09, IEE DOI 10.1109/AMS,2009, 15.

[18] Samsudin, R., Saad, P., and Shabri , A. 2011, River flow time series using least square support vector machines, Hydrol,Earth Syst. Sci, 15, 1835-1852, 2011.

[19] Sihui, D. 2009, Forecast Model of Hydrologic Single, Element Medium and Long-Period Based on Rough Set Theory, IEEE, 978—07695-3735-1/09.

[20] Spate, J.M., Croke, B.F.W., and Jakeman, A.J., 2003, Data Mining in Hydrology, Conference: MODSIM , 2003.

[21] Tapas, K., David, M., Nathan, S., Christine, D., and Angela, Y. 2002, An efficient k-Means Clustering Algorithm: Analysis and Implementation", 0162-8828/02,2002 IEE 24, No.7.

[22] Weilin, L. 2011, Neural network model for hydrological forecasting based on multivariate phase space reconstruction, IEEE, Vol.,2,663-667.

[23] Wiriyarattanakul, S., Auephanwiriyakull, S., Theera, and Umpon, N, 2008, Runoff Forecasting using Fuzzy Support Vector Regression, IEEE, 978-1-4244-2565-5/08.

[24] Wu, C.L., and Chaw, K.W., 2010, Data-driven models for monthly stream flow time series Prediction, Engineering Applications of Artificial Intelligence, Vol. 23, Issue 8, 2010, 1350-1367.

[25] Yuelong, Z., Shijin, L., Dingsheng, W., and Xiaohua ,Z. 2008, A Novel Approach to the Similarity Analysis of Multivibrate Time series and its Application in Hydrological Data mining, 978-07695-3336-0/08, IEE DOI 10.1109/CSSE.2008.1064.