Journal of Hydrology: Regional Studies 4 (2015) 313–332

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

journal homepage: www.elsevier.com/locate/ejrh

Rain gauge network design for flood forecasting using multi-criteria decision analysis and clustering techniques in lower Mahanadi river basin, India

Anil Kumar Kar[a], A.K. Lohani[b,*], N.K. Goel[c], G.P. Roy[a]

[a] Department of Water Resources, Government of Orissa, Bhubaneswar, India
[b] National Institute of Hydrology,Roorkee 247667, India
[c] Department of Hydrology, Indian Institute of Technology, Roorkee, India

A R T I C L E I N F O

Article history:
Received 12 December 2014
Received in revised form 7 June 2015
Accepted 2 July 2015
Available online 25 July 2015

Keywords:
Hall's method
Analytical hierarchical process
Prior network
ANN
Fuzzy logic
NAM model
Key rain gauge network
Flood forecasting

A B S T R A C T

Study region: Mahanadi Basin, India.
Study focus: Flood is one of the most common hydrologic extremes which are frequently experienced in Mahanadi basin, India. During flood times it becomes difficult to collect information from all rain gauges. Therefore, it is important to find out key rain gauge (RG) networks capable of forecasting the flood with desired accuracy. In this paper a procedure for the design of key rain gauge network particularly important for the flood forecasting is discussed and demonstrated through a case study.
New hydrological insights for the region: This study establishes different possible key RG networks using Hall's method, analytical hierarchical process (AHP), self organization map (SOM) and hierarchical clustering (HC) using the characteristics of each rain gauge occupied Thiessen polygon area. Efficiency of the key networks is tested by artificial neural network (ANN), Fuzzy and NAM rainfall-runoff models. Furthermore, flood forecasting has been carried out using the three most effective RG networks which uses only 7 RGs instead of 14 gauges established in the Kantamal sub-catchment, Mahanadi basin. The Fuzzy logic applied on the key RG network derived using AHP has shown the best result for flood forecasting with efficiency of 82.74% for 1-day lead period. This study demonstrates the design procedure of key RG network for effective flood forecasting particularly when there is difficulty in gathering the information from all RGs.

1. Introduction

Information of rainfall is the primary requirement of all flood forecasting models. It is not always possible to gather information from all rain gauges (RGs). The reason could be many. In particular, during flood time there may be chances of failure, breaking, non-recording of RGs, difficulty in transmission of information etc. In large catchments these uncertainties are more prominent. Furthermore it is also mentioned that the climatic changes affect rainfall amounts, rainfall patterns, runoff amounts, and runoff coefficients (Ponce et al., 1997). Research for establishing key RG network is imperative. Kagan (1966) suggested a procedure to compute the error in estimation of aerial rainfall which could be used in estimation of

* *Corresponding author.
E-mail addresses: aklnih@gmail.com, akl_nih@yahoo.co.in (A.K. Lohani).

http://dx.doi.org/10.1016/j.ejrh.2015.07.003

key network density of RGs. Hall (1972) suggested a rational method for determination of key station network. Lohani and Arora (1992) suggested key network stations for flood forecasting purpose. Morin et al. (1979) advocated the use of principal component analysis in conjunction with optimal interpolation for RG network design. Sreedharan and James (1983) used the spatial correlation technique proposed by Kagan for design of RG network. Cheng et al. (2008) have applied a geo-statistical approach for evaluation and augmentation of an existing rain-gauge network in northern Taiwan. Lohani and Arora (1995) compared various precipitation network design methods.

Saaty (1980) has introduced analytical hierarchical process (AHP) for solving the complex decision oriented problems. It can make decisions involving many kind of concerns including planning, setting priorities, selecting the best among a number of alternatives and allocating resources. From its inception, and arising from its concise mathematics and easily obtained input data, the AHP has been of great interest to researchers of many different fields (Triantaphyllou and Mann, 1995; Lin (2006)). Anane et al. (2008) have located and ranked suitable sites for soil aquifer treatment in Jerba Island by integrating a single-objective AHP method into a GIS model. Sinha et al. (2008) have done a multi-parametric approach using AHP and integrates geo-morphological, land cover, topographic and social (population density) parameters to propose a Flood Risk Index for Kosi River. Kevin et al. (2009) has initiated AHP for finding of best management practices in selection and design of storm water schemes. An AHP can effectively deal with both qualitative and quantitative factors in multiple criteria decision environments. AHP is an important decision tool because of its ability to synthesize multi-attributed scenarios and provide diagnostic information, which enables decision makers to better understand the behavioral processes underlying choices.

Acreman and Sinclair (1986) used hierarchical clustering (HC) algorithm for clustering basins for flood frequency in Scotland according to their physical characteristics. Fovell and Fovell (1993) used HC along with principal component analysis (PCA) for identifying climatic regions of the United States based on monthly rainfall and temperature data. Lim and Lye (2003) used hierarchical clustering (average linkage method) in order to delineate homogeneous sub-regions in Sarawak, Malaysia. Jingyi and Hall (2004) used K-mean, Fuzzy C-mean, hierarchical clustering (Ward's method) and Kohonen self organising feature map for clustering the Gan-Ming River basin of China. Kar et al. (2011) used clustering technique for regional flood frequency analysis.

Earlier artificial neural network (ANN) has been applied successfully in several fields of water resources. However the applications in the field of rainfall–runoff modeling and flood forecasting are most popular. The works of Tokar and Markus (2000), Zhang and Govindaraju (2000), Harun et al. (2001), Rajurkar et al. (2002), Sudheer et al. (2002), Riad et al. (2004), Kalteh (2008), Modarres (2009) have shown significance of ANN in rainfall–runoff modeling. In flood forecasting, there is need for models capable of efficiently forecasting water levels and discharge rates. In this regard application of ANN is more effective and the works of Minns and Hall (1996), Campolo et al. (1999, 2003), Imrie et al. (2000), Lekkas et al. (2004), Coulibaly et al. (2000), Dawson and Wilby (2001), Muhamad and Hassan (2005), Mukerjee et al. (2009), Kar and Lohani (2010), Kar et al. (2011), Agarwal et al. (2013) and Lohani and Krishan (2015) emphasized the capability of artificial neural networks over other methods.

Simultaneously, fuzzy logic based approaches (Zadeh, 1965) are also found suitable for hydrological modeling like rainfall–runoff, flood forecasting and risk assessment (Rai et al., 2014). Hundecha et al. (2001) developed fuzzy rule based routines to simulate different processes involved in the generation of runoff from precipitation. These routines were implemented in Neckar river catchment of Germany within a conceptual, modular and semi-distributed HBV model. Casper et al. (2007) framed the runoff model taking soil moisture and rainfall as input through a fuzzy rule based model. Nayak et al. (2005) developed the fuzzy model for forecasting the river flow of Narmada River basin in India. Lohani et al. (2005a, 2014) have applied fuzzy logic in real time flood forecasting of Narmada River basin. Capability of fuzzy logic in hydrologic modeling and forecasting is also demonstrated by Yu and Chen (2005), Valenca and Ludermir (2009) and Lohani et al. (2005b, 2006, 2011, 2012).

The conceptual NAM model is also very popular in rainfall–runoff modeling and thereby very useful for setting the flood forecasting models. Dharmasena (1997) successfully applied Mike-11 package to simulate one-dimensional unsteady flow. He also found conceptual models giving better results especially for rivers subjected to prolonged droughts. Tingsanchali and Gautam (2000) compared two lumped conceptual hydrologic models like tank and NAM with a neural network model applied in two river basins in Thailand. The works of Rabuffetti and Barbero (2005) also showed the application of NAM model.

In this study different possible key RG networks have been designed from the available RG network in the Tel sub-catchment, Mahanadi, India using Hall method, AHP, HC and SOM. Efficiency of the key networks is tested by ANN, Fuzzy and NAM and the best network has been used for flood forecasting. Further, flood forecasting has been carried out with the key RG networks. Although, the best RG network has shown highest efficiency, simultaneously other networks were also tested with certain designated efficiency in order to use them at the time of failure of the best RG network.

2. Study Area

The Mahanadi basin is one of the major basins of eastern part of India. It lies between 80°-30′ to 86°-50′ of East Longitude and 19°-20′ to 23°-35′ of North Latitude. The total catchment area of the basin is 141,569 km^2 comprises major part of two states Chhatisgarh and Orissa. The major reservoir Hirakud with 83400 km^2 catchment lies at central part of the Mahanadi basin.

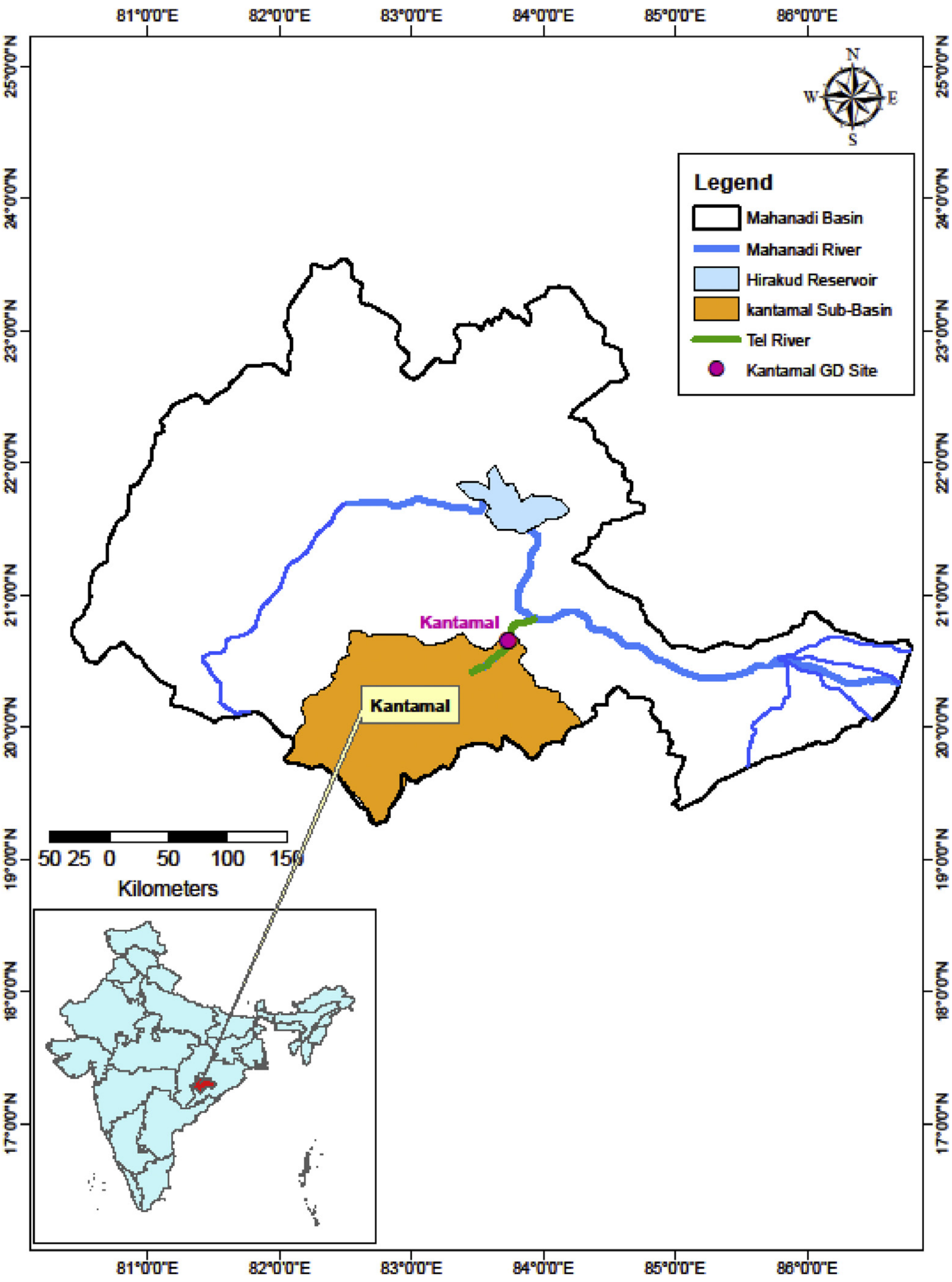

Fig. 1. Location of Mahanadi basin with Kantamal sub-basin.

The downstream part below Hirakud having a catchment of around 50,000 km^2 contributes substantially to flood at delta and is devoid of a sound flood forecasting system. This part has three main tributaries like Jeera, Ong and Tel with catchments 2383, 5128 and 25045 km^2, respectively. Therefore, the contribution from the Tel catchment always remains predominant. Even the flood of 2008 is mainly due to the contribution of this tributary. It has produced a peak discharge of 33762 cumecs during 2008. Keeping this in view, establishment of a flood forecasting model at Kantamal (Catchment of 19600 km^2) upstream of Patharla is attempted in this study. The river Tel joins at Patharla to the main river Mahanadi. The location of Mahanadi basin with Kantamal sub-basin is shown in Fig. 1.

Table 1
Location of RG stations of Kantamal sub-basin with station IDs.

Station ID	Location	Latitude(deg)	Longitude(deg)
KI8	Bhaskel	19.70833	82.13333
KI9	Kurumuli	19.25556	82.82667
M25	Sagada	20.64639	84.00056
M22	Magurbeda	20.78194	83.35833
M16	Goria	20.60556	83.57389
M1	Patora	20.6675	82.44111
M14	Baragaon	20.41111	83.21944
M19	Ichhapur	20.59583	82.59361
M15	Takala	20.25139	82.85222
M18	Chhatikud	19.97222	83.30278
M17A	Burat	20.18694	83.50722
M17	Tulaghat	20.27389	83.57389
M20	Surubali	20.17167	83.77944
R5	Pipalpankha	19.82667	84.33194

3. Data availability

The daily rainfall data of 14 RGs of the study area have been collected for 6 years (2000–2005). The individual RGs have their designated IDs as fixed by concerned department (Department of Water resources, Government of Odisha, India) shown in Table 1. The corresponding daily discharge at Kantamal and daily evaporation data of nearby station at Dasapalla are also collected. The available data is divided into calibration period (2000–2003) and validation period (2004–2005) and only monsoon period (June–October) is considered for the purpose. The physical characteristics of individual Thiessen polygon areas are derived from freely downloadable SRTM data of 90m resolution by using ARCGIS 9.3 software. The hydro-meteorological variables are collected from Department of Water Resources, Government of Orissa.

4. Methodology

In this study our basic aim is to find out key network of RGs (instead of taking information from all) that can be used for making reasonably accurate flood forecasts particularly during the time of distress (when the rainfall data of not all the stations are available due to various reasons). The methodology is basically divided into five parts as:

i) Derivation of storms in different RGs for applying Hall's method.
ii) Derivation of attributes (variables) for application into clustering methods and AHP.
iii) Process to find important (key) gauge networks using prioritizing methods Hall and AHP.
iv) Investigation of possible clusters influencing the model most, using HC and SOM.
v) Test the efficiency of key networks using ANN, Fuzzy and NAM model.
vi) Flood forecasting based on efficient network using ANN and Fuzzy.

The step of methodology is shown in Fig. 2. The flow generation characteristic of each RG is different from each other. The property of each Thiessen area is represented through the physical and hydro-meteorological variables. The variables should be carefully selected and derived as these will be applied for finding key RG networks. In the process to find key RGs two prioritizing methods like Hall, AHP and two clustering methods SOM and HC are adopted. The Hall method forms the network of key RGs considering the storm characteristics only. Whereas, other methods like AHP, SOM, HC are dependent upon the characteristics of the Thiessen area occupied by each RG. The SOM is an unsupervised clustering technique, which is helpful in getting the possible number of divisions. The dendrogram generated by HC method also gives a justification to the decision of SOM. But AHP in this regard makes a ranking of RGs depending upon the influential characteristics of each Thiessen area. Thus key network decision is taken in a multifaceted way and efficiency of each is tested with both soft computing models (ANN, Fuzzy) and conceptual base model (NAM). The brief descriptions of methods adopted are provided blow:

4.1. Derivation of storms

The isolated continuous rainfalls recorded at different RGs are selected as storms over a particular area. These storms are collected over a period, to study the potential RGs susceptible to high rainfall. Prioritization of RGs is being made on the basis of storms occurred applying Hall's method.

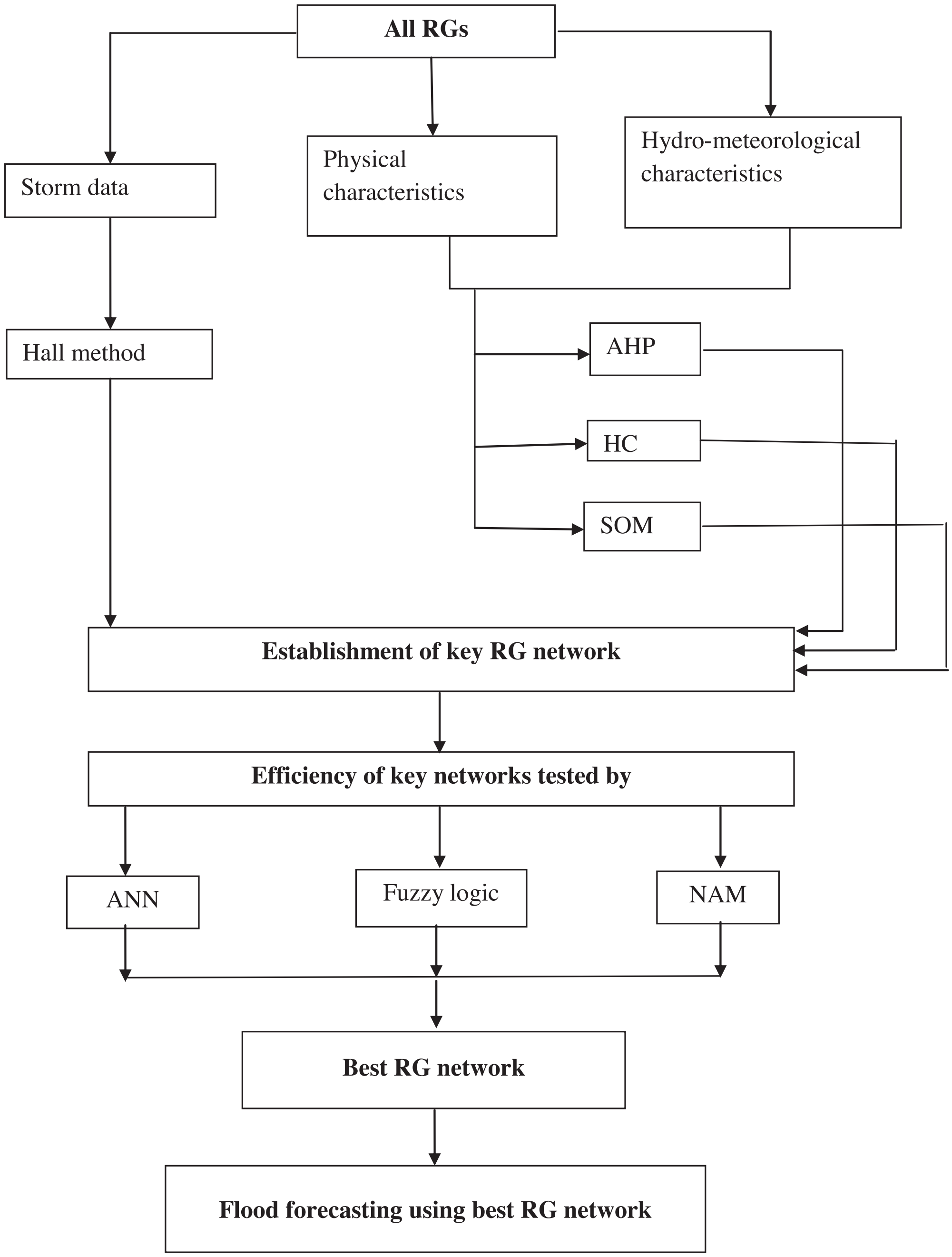

Fig. 2. Flow chart of methodology adopted in this study.

4.2. Derivation of attributes (variables)

The catchment is divided into Thiessen polygons corresponding to the existing RGs and each Thiessen polygon area has to be considered as one unit. The discharge produced by each Thiessen is dependent upon its hydro-meteorological and physical characteristics. Normally those characteristics are to be considered which can be very influential and achievable with less effort. So normally daily average precipitation, maximum 1-day rainfall are some of the important hydro-meteorological characteristics and physical factors like Thiessen weight, average slope, drainage density, longest stream length, soil characteristics, soil moisture content, land use land cover are influential.

In this study daily average rainfall (AP) and maximum 1-day rainfall (1D) are being used as hydro-meteorological variables and in physical attributes Thiessen weight (TW), longest stream of Thiessen area (LS), average slope (SL) and drainage density (DD) are used. The 1D is responsible for generation of floods. A higher 1D can contribute substantially toward formation of peaky floods. When the physical characteristics are in favor of flood formation a higher 1D can generate a higher flood. The AP of an area also reflects the usual rainfall characteristics. The higher AP receiving area saturates the soil more thus converting more rainfall into runoff. Generally a higher AP area receives the maximum 1D and thus generates high flood. Thiessen weight is a relative term of influential area of a RG corresponding to entire catchment area. A large area takes a longer time for draining the runoff to outlet than smaller area. The peak flow per unit area decreases as area increases and the period of surface runoff increases with area Patra, 2002). Longest stream depicts the shape of the catchment toward runoff generation. An elongated catchment with longer stream produces lesser peak than fan shaped area. Slope is an important characteristic of a catchment as it gives an indication of the kinetic energy available for water to move toward the basin outlet, and it has been found to be related to total runoff and base flows (Vogel and Kroll, 1996). Slope may vary frequently within a basin. Subsurface water can only contribute to runoff if a hydraulic gradient exists. The slope of the water table usually conforms to the slope of the land above (Freeze and Cherry, 1979). Accordingly catchments with steep slopes will have high hydraulic gradients, resulting in high base flow rates. Here the average slope of the watershed has been derived from the DEM of the study area. Besides the effect of slope on the hydraulic gradient, steep slopes occur in mountainous and hilly parts of catchments. Some of these hills and mountains have substantial fractures and fissures that store water during the rainy season, and then release it as base flow during the dry season. Drainage density (DD) is derived by dividing the sum of total stream length within a catchment by the catchment area, and is regarded as an important landscape characteristic. It is a measure of how dissected a basin is, and it is expected that DD affects the transformation of rainfall into runoff (Pitlick, 1994; Berger and Entekhabi, 2001). It is expected that fast flow will occur in areas with high DD and steep slope (SL).

4.3. *Process to find key RG network*

4.3.1. *Hall method*

Hall (1972) suggested a rational method for determination of key station network using the equation

$$P_a = C + A_1X_1 + A_2X_2 + A_3X_3 + A_4X_4 \ldots + A_nX_n \tag{1}$$

Where, P_a is the rainfall to be estimated from observed record at selected station X_1,X_2,X_3...X_n.A_1,A_2,A_3, ...A_n are regression coefficients C being a constant known intercept.

In order to establish key station network correlation coefficient between average of the storm rainfall and individual station rainfall are found. The correlation coefficients thus obtained are arranged in a descending order and the highest correlation coefficient station is the first key station and its data is removed from the set. The next set is chosen in the same way and highest correlation coefficient bearer being the second key station and the process continues.

Then a key station network is getting investigated by finding the multiple correlations co-efficient of individual stations with that of average storm of the group. Gradually the stations added to the key station network, the total amount of variance at that stage is determined. With the addition of a RG to a network the multiple correlation coefficient increases and sum of the squares of deviation decreases till a stage is reached when improvement in either the multiple correlation coefficient or sum of the squares of deviation will be negligible. The corresponding number of RGs at this stage is taken as the representative network for the purpose of determining aerial estimate of rainfall.

4.3.2. *AHP*

The analytic hierarchy process (AHP) is a theory of measurement through pairwise comparisons and relies on the judgments of experts to derive priority scales (Saaty 2008). Although its foundation lies in complex matrix manipulation, it is readily implemented and can be used without a great deal of knowledge about multi-criteria decision making. In AHP a goal can be achieved in four steps, like

– Pairwise comparison of alternatives
– Extraction of priority vectors
– Finding consistency of pairwise judgments
– Ranking the priority alternatives

4.3.2.1. Pairwise comparison of alternatives. A matrix is framed containing the criteria/alternative with different choices. These alternatives are pre-selected for testing for a particular type of problem. This matrix remains the basis for evaluating different alternatives for achieving various selection criteria. In our study, matrix between RGs and property of Thiessen area is to be framed. Through this matrix both different choices of RGs as well as choices of different criteria/alternative properties of Thiessen area are compared. This comparison of matrices is dynamic and can be adjusted on separate applications. A pairwise comparison matrix 'A' has to be formed, where the number in i^{th} row and j^{th} column gives the relative importance of O_i (objective) as compared with O_j (objective). The AHP conditional rules are shown in Table 2.

4.3.2.2. Extraction of priority vectors. In this the right eigen vector is decided for both choices (RGs) and alternatives (properties of Thiessen area). It is a collection of RGs ranking vectors (one for each alternative) and a single vector that ranks the alternatives.

Table 2
AHP rules.

a_{ij}(conditional numbers) =	Conditional rules
1	Two objectives are equal in importance.
3	If o_i is moderately more important than o_j .
5	If o_i is strongly more important than o_j .
7	If o_i is very strongly more important than o_j .
9	If $o_i$is extremely more important than o_j .
1.1 to1.9	If o_i and o_j are very close.
2,4,6,8	Intermediate judgment values
Reciprocals like 1/3,1/5	$o_j$is moderately more important than o_i and so on for other reciprocals.

4.3.2.3. Finding consistency of pairwise judgments. In order to prove the strength of our assumption there should be some authentic check and that is in the form of consistency check. Whatever assumptions are taken for alternatives initially that must hold good for all choices otherwise it should be modified. The consistency index (CI) is as per Eq. (2). Where, $\lambda_{max}$is the principal eigen value and n is the total number of activities. The consistency ratio (CR) is to be measured for assumptions of alternatives to be true which is the ratio of consistency index to random index (RI). The value of RI is dependent on the size of matrix getting used (for n = 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15; RI = 0.0, 0.0, 0.58, 0.90, 1.12, 1.24, 1.32, 1.41, 1.45, 1.49, 1.51, 1.53, 1.56, 1.57 and 1.59, respectively). The value of CR should be within 0.1 (Saaty, 1980) for allowing the assumptions.

$$\mathrm{CI} = (\lambda_{\max} - \mathrm{n})\frac{1}{\mathrm{n}-1} \tag{2}$$

$$\mathrm{CR} = \frac{\mathrm{CI}}{\mathrm{RI}} \tag{3}$$

4.3.2.4. Ranking the priority alternatives. In this two level decision matrix we have to decide the priority vector for criteria. The priority vectors are arranged in the matrix shown below. Simultaneously, the choices are again to be put in the same process of selection and priority vectors of choices are to be made against each criterion. Therefore, again a matrix of priority of vectors of choices is made as shown in Eq. (4). For final evaluation of each choice the matrix multiplication of transposed matrix of priority vector of criteria are done with priority vector of choices (Eq. (5)). In our study, there are 6 criteria (alternatives) and each criterion has 14 choices (Eq. (6)). The final ranking of choices is presented in Eqs. (7)–(9)

$$\left\{\begin{matrix} A_1 \\ A_2 \\ . \\ A_6 \end{matrix}\right\} \tag{4}$$

$$[A_1 A_2 \ldots A_6] \tag{5}$$

$$\left\{\begin{matrix} a_{1,1} & a_{2,1} & \ldots & a_{14,1} \\ b_{1,1} & b_{2,1} & \ldots & b_{14,1} \\ \vdots & & & \\ n_{1,1} & n_{2,1} & \ldots & n_{14,1} \end{matrix}\right\} \tag{6}$$

$$\text{Rankofchoice}, a_1 = A_1^T a_{1,1} + A_2^T a_{2,1} \ldots + A_6^T a_{14,1} \tag{7}$$

$$\text{Rankofchoice}, b_1 = A_1^T b_{1,1} + A_2^T b_{2,1} \ldots + A_6^T b_{14,1} \tag{8}$$

$$\text{Rankofchoice}, n_1 = A_1^T n_{1,1} + A_2^T n_{2,1} \ldots + A_6^T n_{14,1} \tag{9}$$

The overall consistency of the process can also be checked by using the Eq. (10) with permissible value of $\bar{\mathrm{CR}} \geq 0.1$.

$$\bar{\mathrm{CR}} = \frac{\sum_i \mathrm{w_i CI_i}}{\sum_i \mathrm{w_i RI_i}} \tag{10}$$

4.4. Application of clustering methods

4.4.1. Hierarchical clustering

In cluster analysis a group of elements segmented into closely related groups using the number of collected attributes of the elements. In an hierarchic classification the data are not partitioned into classes in one step. Rather they are first separated into few broad classes each of which further divided into smaller classes and each of these further partitioned, and so on until terminal classes are generated which are not further subdivided (Everitt, 1980). Formation of a group depends on the degree of similarity (or dissimilarity) among the elements. In hierarchical clustering the data are first separated into few major classes and each of these further subdivided. These small classes portioned to form a tree structure called dendrogram.

In this study HC has been used to check the possible distribution of RGs into groups as per individual characteristics. The suitable groups are then to be formed after visualizing the resultant dendrogram.

4.4.2. Self organization map

The self organization map is invented to reduce the data dimension using self organizing neural networks (Kohonen, 1982). The internal structure of the data is well visualized through this. The Kohonen map based data- clustering technique is applied to show how multi-dimensional datasets can be reduced to 2-Dimensional feature maps, manifesting clusters of similar data items (Kiang et al., 1997).

SOM is a two layered structure containing an input layer and an output layer. Each neuron is connected to the input data with their own weights. In the case of SOM weights will be equivalent to the normalized attributes representing the centroid coordinates of each cluster. This feature of the SOM is very different than the feed forward fully connected ANN, where weights do not relate directly to the input data. The output layer is usually one or two-dimensional and its nodes may be connected to each other. The input neurons transfer the object features to the output neurons. The learning procedure in a Kohonen Map is unsupervised competitive learning. Only the winning node and its neighbors are updated during the learning. Weights w_{ij} are updated using following formula in Eq. (11).

$$\mathrm{w_{ij}(new) = w_{ij}(old)} + \propto [\mathrm{x_i\text{-}w_{ij}(old)}] \tag{11}$$

Where, x_i is the i^{th} input signal, w_{ji} is the weight of the connection from node i to node j and $\propto$ is the learning rate. The winning node is determined by a similarity measure, which can be Euclidean distance measure or the dot product of two vectors. The Euclidean distance (D_j) that is mostly used for similarity measure is calculated as per Eq. (12).

$$\mathrm{D_j} = \sqrt{\sum_{\mathrm{i=1}}^{\mathrm{n}} (\mathrm{x_i - w_{ij}})^2} \tag{12}$$

In our study it is applied

i) to test the possible number of groups in the RG datasets
ii) divide the RGs into different groups to check the efficiency of each group.

4.5. Testing the efficiency of key networks

4.5.1. ANN model

A 3-layered (n-k-o) ANN model is selected to test the output of different model. Where n is input neurons, k hidden neurons and o output neuron. The MATLAB software has been used to solve this network. The multi layered feed forward network is used with back propagation error modeling. The inputs are daily rainfall data of different RGs and evaporation data is one of the nearby sites. The data is normalized to 0–1 scale. The optimum number of neurons in the hidden layer was identified using a trial and error procedure by varying the number of neurons in the hidden layer. In all cases 'tansig' neurons are used in first layer, 'purelin' in second layer and 'trainbr' remains the training function. The weights and biases of the networks are adjusted using gradient descent with momentum weight and bias learning function. Number of trials has been made to get a consistent result.

4.5.2. Fuzzy model

The linguistic fuzzy model maps the characteristic of input data to input membership functions, input membership function to rules, rules to a set of output characteristics to output membership function and output membership function to a single valued output or a decision associated with the output (Jang et al., 2002). Whereas, a fuzzy rule based model suitable for the approximation of many systems and functions is the TS fuzzy model in which the consequents are expressed as (crisp) function of the input variables (Takagi and Sugeno, 1985). It is defined as:

R_i: IF m_i is A_{i1} AND . . .AND IF m_n is A_{in}

$$\text{THEN}\; P_i = a_{i1}m_1 + a_{i2}m_2 + \ldots a_{\mathrm{in}}m_n + b_i \tag{13}$$

Where R_i (i = 1,r) indicates a set of r rules derived by partitioning the input clustering approach proposed by Chiu (1994). m_1,m_2. . .m_n are the input variables in n dimensional input vector and $p_i \in \hat{A}$ is the consequent of the i^{th} rule. In the consequent,a_i is the parameter vector and b_i is the scalar offset. A_{in} is the (multivariate) antecedent fuzzy set of the i^{th} rule.

The rules framed for operation may be set manually but rules based on clustering make it simpler. The data driven approach based on subtractive clustering has shown promising results in various hydrological modeling application (Lohani et al., 2005a,b). The purpose of subtractive clustering is to identify natural grouping of the data from a large dataset and finally to produce a concise representation of a system behavior (Lohani et al., 2006, 2007a,b).

Table 3
Key RGs as per priority (Hall Method).

Normal sequence	Prioritized Stations	Correlations
R5	M1	0.901
KI8	M17A	0.876
KI9	M22	0.884
M25	M15	0.872
M22	M18	0.865
M16	M17	0.872
M1	M20	0.818
M14	M16	0.806
M19	M25	0.791
M15	M14	0.782
M18	M19	0.791
M17A	KI8	0.802
M17	KI9	0.882
M20	R5	0.491

4.5.3. NAM

NAM is the abbreviation of the Danish "Nedbør–Afstrømnings-Model", meaning precipitation-runoff model, part of the rainfall–runoff (RR) module of the MIKE 11 river modeling system. This model was originally developed by the Department of Hydrodynamics and Water Resources at the Technical University of Denmark. It is a deterministic, lumped and conceptual rainfall-runoff model which simulates the rainfall–runoff processes occurring at the catchment scale. This model is well adopted in different climatic zones of the world

NAM also uses auto calibration optimizing all 9 parameters automatically. The four different objectives like water balance, overall hydrograph shape, peak flows and low flows remains the basis of auto calibration.

A mathematical hydrological model like NAM is a set of linked mathematical statements describing, in a simplified quantitative form, the behavior of the land phase of the hydrological cycle. NAM represents various components of the rainfall-runoff process by continuously accounting for the water content in four different and mutually interrelated storages. Each storage represent different physical elements of the catchment. NAM can be used either for continuous hydrological modeling over a range of flows or for simulating single events.

5. Results and discussions

The study area has 14 RGs and the corresponding discharge is measured at Kantamal G&D site. The evaporation data of nearby site is also considered for application into study area.

5.1. Selection of key network station

For the selection of key network stations Hall, AHP, SOM and HC methods have been applied. The results of these methods are discussed below:

5.1.1. Hall method

From the daily rainfall data of six years, in total 29 storms are identified. The correlation coefficient between average of the storm rainfall and individual station rainfall are found and arranged in a descending order. The highest correlation coefficient station is the first key station and its data is removed from the set. The next set is chosen in the same way and highest correlation coefficient bearer being the second key station and the process continues (Table 3). It is seen that the station M1 gives highest correlation and R5 the lowest.

Then the key station network was investigated by finding the multiple correlation coefficients (MCC) of individual stations with that of average storm of the group. Gradually the stations added to the key station network, the total amount of variance at that stage is determined (Table 4). A plot of MCC and RGs has also been made to check the substantial improvements after addition of RGs into the network (Fig. 3a). It is depicted from Fig. 3(a) that, the MCC is rising significantly upto addition of M17 but after the addition of M20 the MCC drops and then again it increases with the addition of M16. So M20 is a wrong fit here and after removing it the graph shows continuous growth (Fig. 3b). This growth is also not substantial after addition of M16. Thus a key network with 7 stations (M1, M17A, M22, M15, M18, M17 and M16) is finalized with MCC of 0.996 (Network HM). Addition of further stations is not very vital after achieving this much of MCC.

5.1.2. AHP

AHP has been applied according to Fig. 4. In order to achieve the target i.e. to form a key network of RGs, the first level (Criteria level) is to be carefully selected. The criteria have been chosen in such way that the selected attributes should be easily available to practicing hydrologists without much effort. Here, in first level 6 criteria like DD, 1D, AP, SL, TW and LS have been considered. Further, in the second level there are choices of 14 RGs for each criterion. The DEM of the study area

Table 4
Multiple correlation coefficients of individual RGs with average storms of the sub-basin.

	M1	M17A	M22	M15	M18	M17	M20	M16	M25	M14	M19	KI8	KI9	R5	Intercept	MCC
1	0.843														37.34	0.949
2	0.492	0.286													32.85	0.971
3	0.226	0.256	0.392												25.91	0.985
4	0.104	0.212	0.383	0.223											24.48	0.988
5	0.077	0.163	0.301	0.226	0.162										17.52	0.992
6	0.437	0.119	0.259	0.174	0.186	0.085									17.63	0.994
7	0.043	0.122	0.271	0.177	0.184	0.084	-0.015								17.9	0.993
8	0.137	0.098	0.353	0.076	0.077	0.063	-0.073	0.347							15.01	0.998
9	0.127	0.116	0.224	0.131	0.092	0.059	-0.075	0.236	0.135						15.4	0.998
10	0.171	0.117	0.191	0.079	0.045	0.068	-0.053	0.192	0.114	0.078					15.4	0.998
11	0.163	0.12	0.183	0.077	0.044	0.066	0.05	0.18	0.119	0.081	0.017				15.74	0.998
12	0.087	0.019	0.134	0.104	0.083	0.096	0.077	0.154	0.087	0.069	-0.015	0.122			6	0.999
13	0.065	0.064	0.112	0.077	0.059	0.077	0.049	0.139	0.069	0.065	0.071	0.091	0.059		5.72	0.999
14	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0.071	0	1

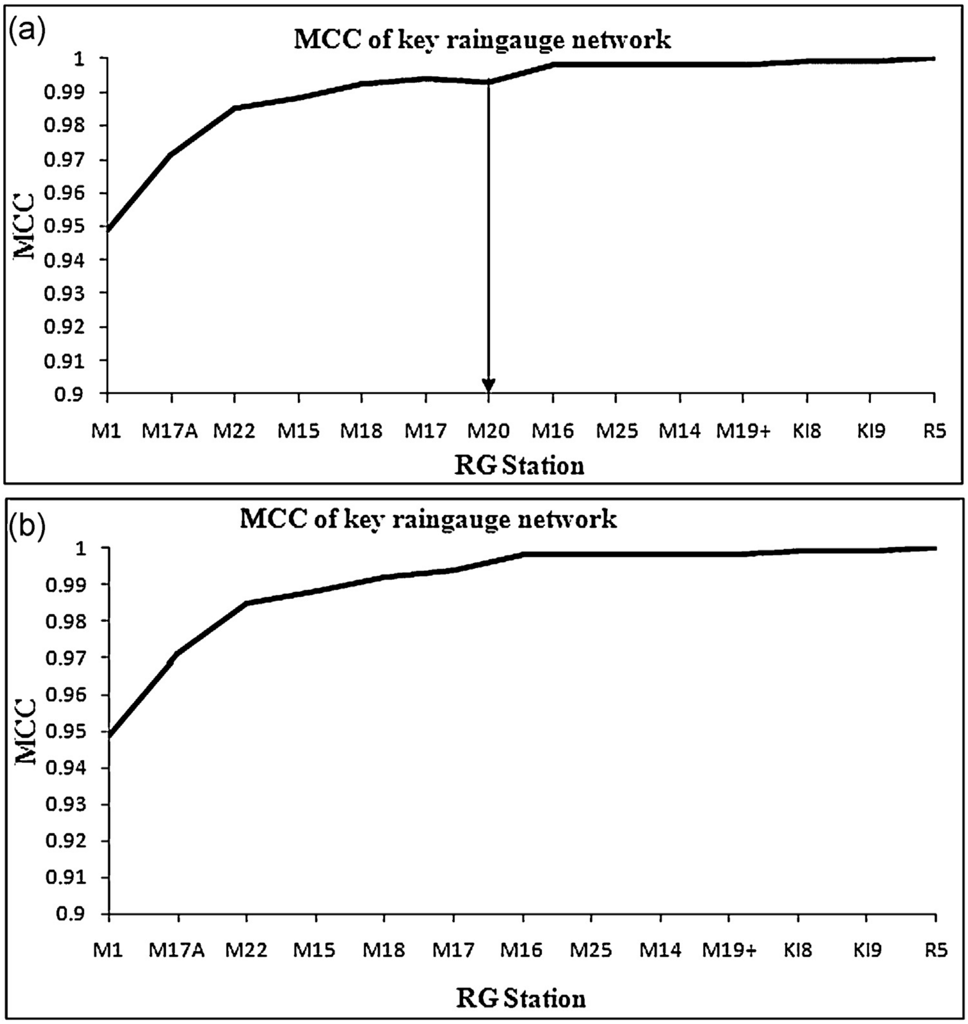

Fig. 3. (a) Increase in MCC for formation of key RGs network (all stations). (b) Increase in MCC for formation of key RGs network (after removing M20).

has been prepared from the freely available SRTM data. Thiessen areas of all 14 RGs have been computed from the Thiessen polygon map of the study area (Fig. 5). The hydro-meteorological characteristics like 1D and AP have been derived from the daily rainfall data. The physical characteristics have been obtained using ARCGIS. Further, on the basis of the characteristics of individual RGs AHP has been applied.

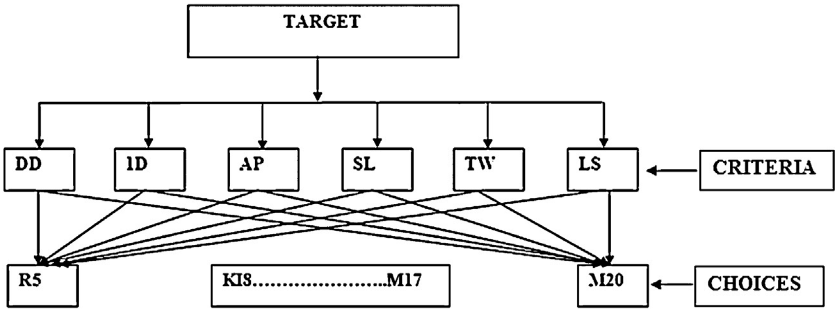

Fig. 4. AHP diagram of our study.

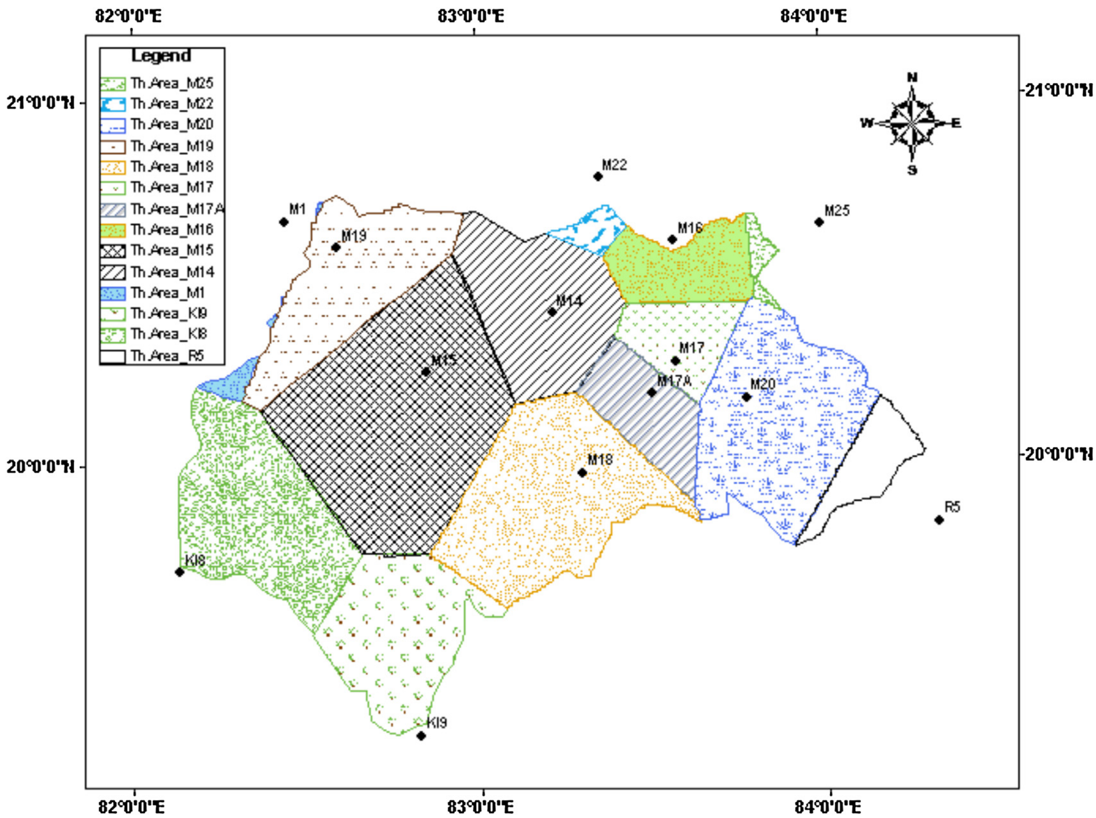

Fig. 5. Thiessen polygons of 14 stations of study area.

Table 5
Result of PCA.

Principal components	Eigen value	Percentage variance
1	2.587	43.121
2	1.602	26.692
3	1.264	21.07
4	0.266	4.431
5	0.188	3.14
6	0.093	1.546

A pair wise comparison has been made among the available 6 criteria and importance of each criterion over other has been measured in a 1–9 point scale as shown in the judgment matrix table (Table 7). In order to further strengthen the judgment, a principal component analysis has been applied over the six available criteria. It is seen from Table 5 that considering 3 PCs more than 90% variance is restored. The loadings of the first 3 PCs (having eigen value more than 1) in Table 6 say the first PC is governed by DD and RL, second by AP and 1D and third by SL. The TW and LS never dominates either PC. In runoff

Table 6
Loadings of first 3 PCs.

Choice	PC1	PC2	PC3
SL	−0.3	0.3482	−0.6244
ID	0.3516	0.5635	0.2085
DD	0.5192	−0.0948	0.3757
AP	0.2363	0.6701	−0.0991
TW	0.4447	−0.2859	−0.4915
LS	0.5128	−0.1471	−0.4173

Table 7
The judgment matrix for criteria.

Choice	LS	TW	SL	ID	DD	AP	
LS	1	0.333	0.2	0.2	0.11	0.2	
TW	3	1	0.333	0.333	0.2	0.5	
SL	4	3	1	0.333	0.333	0.5	
ID	5	3	3	1	0.5	2	
DD	9	5	3	2	1	2	
AP	5	2	2	0.5	0.5	1	
Sum	27	14.33	9.53	4.37	2.64	6.2	Total = 64.08

Table 8
Normalized judgment matrix.

Choice	LS	TW	SL	ID	DD	AP	Sum	Priority vector (%)
LS	0.037	0.0233	0.021	0.0458	0.042	0.0323	0.2013	3.3558
TW	0.1111	0.0698	0.035	0.0763	0.0756	0.0806	0.4485	7.4742
SL	0.1481	0.2093	0.1049	0.0763	0.1261	0.0806	0.7454	12.423
1D	0.1852	0.2093	0.3147	0.229	0.1891	0.3226	1.4498	24.1639
DD	0.3333	0.3488	0.3147	0.458	0.3782	0.3226	2.1556	35.9267
AP	0.1852	0.1395	0.2098	0.1145	0.1891	0.1613	0.9994	16.6563
Sum	1	1	1	1	1	1	6	100

generation all DD, AP, 1D and SL has important roles but in consideration to this particular catchment, the physical properties of each variable and result of PCA the judgment matrix is finalized as per these considerations given below. As per PC1, the loadings of DD and LS are higher. But LS is not a more powerful variable than DD. A longer LS can lead to more infiltration and other losses, longer time to peak, so an attenuated peak is received. Whereas DD allows more percentage of rain water getting converted to runoff and lower travel times, so lower loss of rain water and higher peaks are available. Thus DD is considered to be a more powerful variable than LS and as DD belongs to PC1 its influence is considered as more important than other variables. In PC2 both AP and 1D have more or less the same loadings. AP influences the runoff in the long term and 1D instantly. The area having higher AP has saturated the soil more thus more runoff is generated whereas higher 1D remains the immediate cause for higher peaks. In PC3, SL is the highest loading variable. Although it is the most influencing variable for runoff generation here in site specific case it is considered next to DD, AP and 1D. The other choices TW and SL are not so predominant in physical properties also, so is considered next to SL.

The judgments for this study are made in this way:-

- Thiessen weight (TW) is moderately more important than longest stream length (LS).
- Slope (SL) is between moderately and strongly more important to longest stream length (LS) and moderately more important to Thiessen weight (TW).
- Daily maximum rainfall (1D) is strongly more important to longest stream length (LS), moderately more important than Thiessen weight (TW) and moderately more important to slope (SL).
- Drainage density (DD) is extremely more important than longest stream length (LS), strongly more important to Thiessen weight (TW). Moderately important to slope (SL) and between equal and moderately important to maximum 1-day rainfall (1D).
- Daily average rainfall (AP) is very strongly more important to longest stream (LS) and between equal and moderately important to Thiessen weight (TW) and slope (SL), equal or moderately less important to daily maximum rainfall (1D) and drainage density (DD).
- The upper part of the diagonal in the judgment matrix is the reverse of the bottom part and vice versa.

The judgment matrix in Table 7 is then normalized by dividing each value by sum of that column and the principal eigen vector (priority vector) is obtained by adding the normalized values along each row (Table 8), which shows that DD is highly influential criteria followed by 1D, AP, SL, TW and LS. The physical influence of each variable toward runoff generation is given by the importance for establishing the comparison matrix.

Table 9
Overall ranking of choices.

	DD	1D	AP	SL	TW	LS	Priority vector (%)
Weight→↓choices	0.3433	0.2263	0.17167	0.1431	0.0838	0.0319	
R5	1.77	1.77	2.19	20.38	2.06	2.19	4.54
KI8	14.19	2.69	4.39	3.11	11.15	8.54	7.89
KI9	11.1	17.04	17.09	3.11	6.21	4.38	11.7
M25	1.76	2.17	1.78	14.08	1.78	1.77	3.62
M22	4.36	8.54	3.1	1.79	2.19	2.19	4.47
M16	20.16	2.19	1.78	2.2	3.36	3.09	8.42
M1	3.07	6.2	6.22	4.41	1.78	1.77	4.36
M14	6.18	4.38	4.39	1.87	8.28	13.98	5.27
M19	2.05	2.19	2.19	4.41	4.38	4.38	2.71
M15	8.51	3.09	3.09	2.2	20.25	20.24	6.81
M18	3.34	11.14	11.17	11.22	17.06	6.2	8.82
M17A	16.97	20.24	14.02	6.24	4.38	11.14	14.43
M17	2.18	4.38	20.02	7.8	3.13	3.09	6.65
M20	4.36	13.98	8.57	17.16	13.99	17.04	10.3

Table 10
Prioritisation of RGs as per Hall's method and AHP.

Sl. No.	Station name (normal sequence)	Prioritized Stations	
		Hall method	AHP
1	R5	M1	M17A
2	KI8	M17A	KI9
3	KI9	M22	M20
4	M25	M15	M18
5	M22	M18	M16
6	M16	M17	KI8
7	M1	M20	M15
8	M14	M16	M17
9	M19	M25	M14
10	M15	M14	R5
11	M18	M19	M22
12	M17A	KI8	M1
13	M17	KI9	M25
14	M20	R5	M19

The pairwise comparison of initial assumptions of the physical and hydro-meteorological criteria has been checked for its consistency. The λ_{max} which is the sum of the product of sum of each column of judgment matrix and its corresponding priority vector comes as 6.386 and CI = 0.077. For $n = 6$ the value of RI is 1.24 and therefore CR = 0.077/1.24, which comes as $0.062 \leq 0.1$. So the assumptions are consistent.

The same steps again continued for second level of hierarchy. Here there are 14 choices against each criterion. So ranking of choices are done against each criterion. For putting 14 RGs in a 1–9 scale we grouped them in to 10 classes and judged each of them according to their rank with respect to individual criterion. Thus 6 such priority vectors are made available for individual choices. Then the final priority vectors are obtained by multiplying the value of each choice with corresponding weights of criteria obtained from Table 9.

An overall consistency is again checked for justifying all the assumptions made earlier. The $\bar{CR}$ value obtained using the equation 10 is 0.0569 which is ≤ 0.1. On the basis of the final priority vector ranking of individual RGs has been performed and shown in Table 10 against that and compared with the Hall method.

5.1.3. SOM

Further another unsupervised method Kohonen self organization map is also trialed with same datasets in order to find the possible number of groups and RGs constituting each group. The NNclust software has been used in this regard. Several combination of learning parameter, sigma for Gaussian neighborhood and training cycles are considered. The network gave consistent result at learning parameter 0.35 to 0.1, Gaussian neighborhood 30–1% and training cycle of 200. The results (Fig. 6) illustrates that there are two possible clusters separating 14 RGs into two groups of 7 RGs in each. Allotments of RGs in two different clusters are: Cluster 1 includes KI8, KI9, M14, M15, M18, M17A, M20 and cluster 2 includes R5, M25, M22, M16, M1, M19, M17. Taking each group as a possible network, two possible networks have been obtained with KI8, KI9, M14, M15, M18, M17A, M20 as SOM1 and R5, M25, M22, M16, M1, M19, M17 as SOM2.

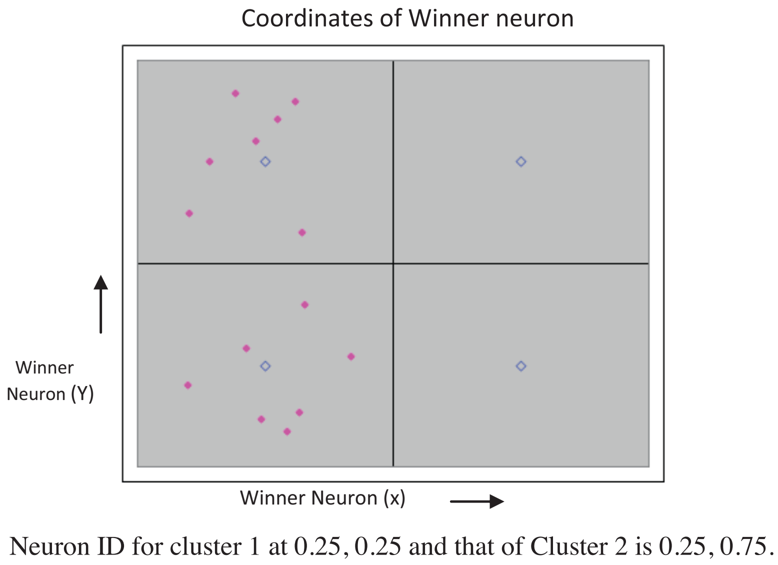

Fig. 6. RGs distributed in two clusters through Kohonen self organization map.

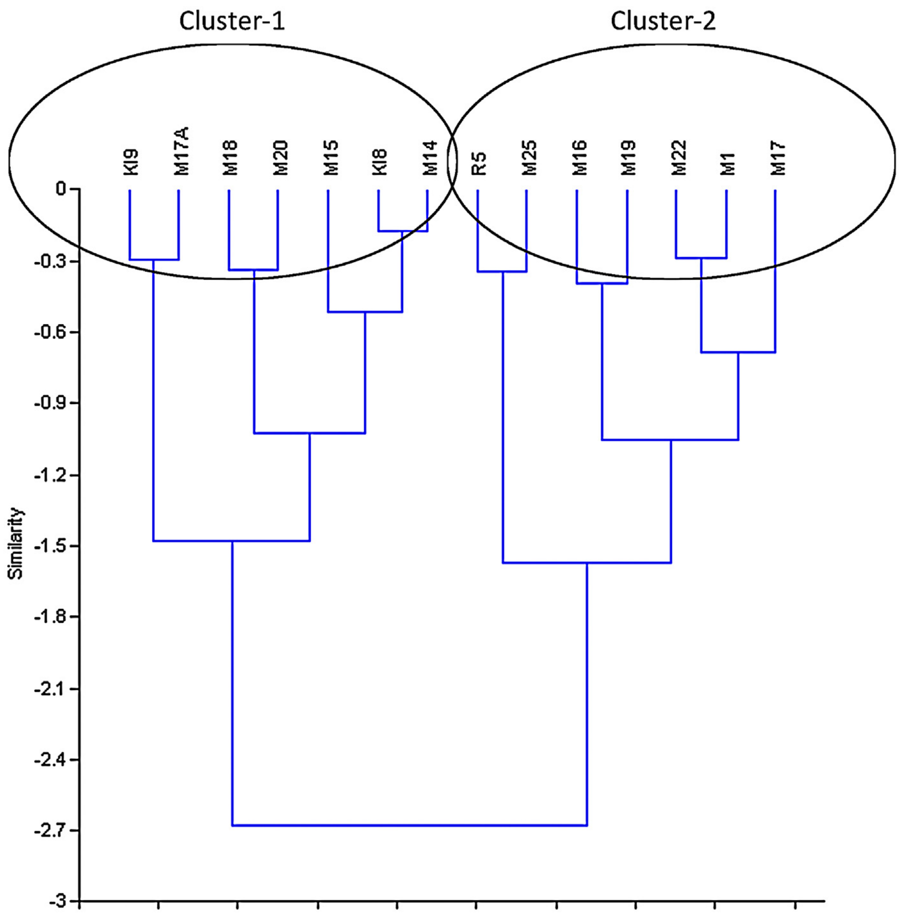

Fig. 7. The dendrogram of HC dividing 14 RGs in two groups.

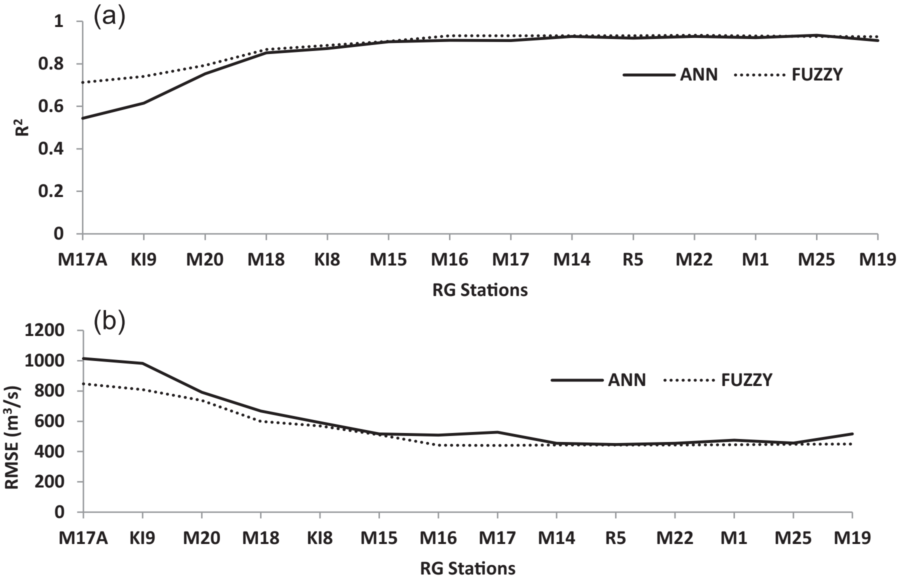

Fig. 8. (a) Comparison of R2 in AHP using ANN and Fuzzy. (b) Comparison of RMSE in AHP using ANN and Fuzzy. (c) Comparison of Efficiency in AHP using ANN and Fuzzy.

Table 11
Key rain gauge networks by all four methods.

Priority Methods	Networks	Network name	Rain gauge names
Hall	1	HM	M1,M17A,M22,M15,M18,M17,M16
AHP	1	AHP	M17A,KI9,M20,M18,M16,KI8,M15
SOM	1	SOM1	KI8,KI9,M14,M15,M18,M17A,M20
	2	SOM2	R5,M25,M22,M16,M1,M19,M17
HC	1	HC1	KI9, M17A, M18, M20,M15, KI8,M14
	2	HC2	R5,M25, M16, M19, M22, M1, M17

5.1.4. HC

The same variables of 14 RGs have been considered for hierarchical clustering. It is depicted from the dendrogram (Fig. 7) that all the available 14 RGs can be separated into 2 clusters. Further each cluster has been taken as one model (HC1, HC2) in our study. The result of SOM and HC are same.

5.2. Performance measure of key RG networks

Two prioritising methods (Hall and AHP) and two clustering methods (HC and SOM) have been applied in this study. The key RG networks from AHP and Hall methods are tested using ANN and Fuzzy techniques. The performance criteria are fixed as r^2 (coefficient of determination), efficiency (Nash–Sutcliffe criterion) and RMSE (m^3/s) are tested for both networks adding one station each time. It is observed that a major epresentation is achieved with 7 RGs in AHP network, after that whatever growth in performance is achieved by adding further RGs is marginal (Figs. 8 (a–c) and 9 (a–c) . The r^2, RMSE and efficiency of AHP network with 7 RGs are 0.9319, 441.107 and 0.8618 and that of Hall method with same 7 RGs is 0.9021, 521.8355 and 0.8123. The result of AHP network shows better result than Hall network. The fuzzy logic applied performance measures remain higher than that of ANN results.

The two clustering methods HC and SOM also show presence of 7 RGs in each cluster. So for comparative study 7 best RGs are chosen from two network models and 2 each from clustering methods. Thus a total of 6 best key network models (1 from Hall method, 1 from AHP, and 2 each from SOM and HC) have been considered and tested for their efficiency and are presented in Table 11.

The models chosen are tested for rainfall–runoff modeling using ANN, Fuzzy logic and conceptual NAM model. The input data contains 6 years of daily rainfall, evaporation and discharge data divided into calibration and validation periods. The same performance criteria are used as applied earlier for ANN, Fuzzy and NAM models.

First of all a MLFF ANN network has been attempted with varying number of hidden layers, hidden neurons and iterations. Only one hidden layer has been fixed for all models in order to avoid model complexity. For different models it has been observed that the hidden neurons vary from 2 to 7 in numbers, learning rate from 0.4 to 0.6, momentum constant from 0.7 to 0.9 and epochs from 50 to 150 with increment of 20. The best performance measures for different models are collected

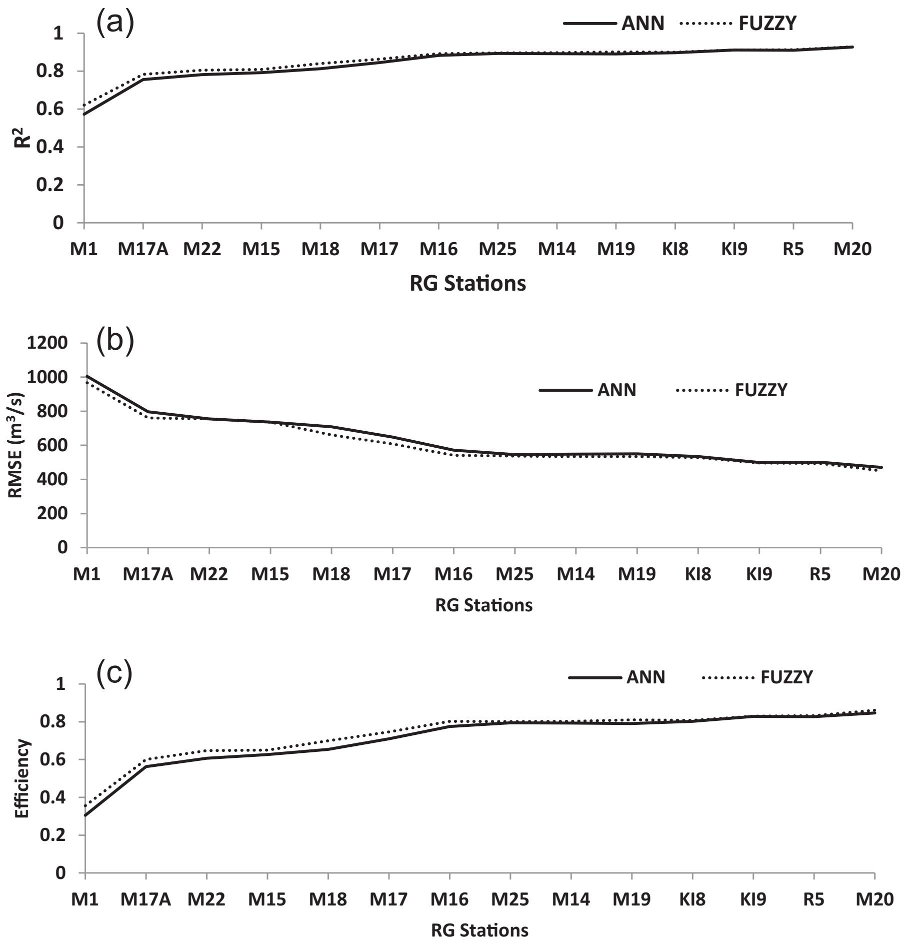

Fig. 9. (a) Comparison of R2 in Hall's method using ANN and Fuzzy. (b) Comparison of RMSE in Hall's method using ANN and Fuzzy. (c) Comparison of Efficiency in Hall's method using ANN and Fuzzy.

and shown in Table 12. For Fuzzy logic approach the Takagi–Sugeno algorithm (Lohani et al., 2007a,b) which uses Gaussian membership function has been applied. The subtractive cluster radius has been varied from 0.1 to 0.7 with an increment of 0.05 in each trial. The performance measures are presented in Table 12.

Similarly in case of NAM model the initial parameters have been derived by auto-calibration. Further, these parameters have been fine tuned using the available basin information. The parameters derived for best calibration results have been used for the validation series. Finally, the performance measures have been derived in the same way as in the case of ANN and Fuzzy approach and presented in Table 12.

It is revealed from the analysis that RG networks established by AHP perform better for all the models derived using ANN, Fuzzy and NAM. However the RG network established by Hall method also performs better but remains inferior to AHP. When comparing the results obtained from all the models (Table 12), the AHP network with Fuzzy R–R model has been observed as the best model with 86.18% efficiency in validation. Similarly, the same AHP network provides the lowest RMSE of 441.107m^3/s.

5.3. Best networks

The best combination has been obtained as AHP which gives coefficient of determination (r^2) in calibration and validation as 0.946 and 0.893, RMSE 486.983, 570.6169 and efficiency as 0.8439, 0.7743 as per ANN. While, the AHP derived fuzzy R–R model gives coefficient of determination (r^2) 0.9243, 0.9319, RMSE 470.461, 441.107, efficiency 0.8544, 0.8618 and the NAM model gives coefficient of determination (r^2) 0.830, 0.602, RMSE 616.4774, 754.7571 and efficiency as 0.830 and 0.601. Comparing the performance criteria of other models the SOM1 or HC1 model is the second best and HM is the third best network model. So in absence of the first network second and third can be attempted to make a reasonable forecast. As per the performance criteria 3 best networks are detailed in Table 13.

Table 12
Performance measure of different networks.

ANN model							
	Coefficient of Determination (r^2)			RMSE (m^3/s)		Efficiency	
	Network	calibration	validation	calibration	validation	calibration	validation
Hall	1	0.925	0.883	528.069	571.716	0.8556	0.774
AHP	1	0.946	0.893	486.983	570.617	0.844	0.774
HC	1	0.9251	0.862	567.534	615.797	0.856	0.738
	2	0.892	0.693	675.819	875.795	0.795	0.471
SOM	1	0.9251	0.862	567.534	615.797	0.856	0.738
	2	0.892	0.6931	675.819	875.795	0.795	0.471
Fuzzy model							
Hall	1	0.903	0.902	539.944	521.835	0.816	0.812
AHP	1	0.9243	0.9319	470.461	441.107	0.8544	0.8618
HC	1	0.904	0.9054	537.938	511.78	0.8174	0.819
	2	0.841	0.7396	808.815	815.188	0.7073	0.542
SOM	1	0.904	0.905	537.938	511.78	0.8174	0.819
	2	0.841	0.739	808.815	815.188	0.7073	0.542
NAM							
Hall	1	0.753	0.611	884.135	893.195	0.65	0.541
AHP	1	0.83	0.602	616.477	754.757	0.83	0.601
HC	1	0.744	0.441	756.87	892.64	0.744	0.442
	2	0.614	0.546	928.06	835.95	0.615	0.511
SOM	1	0.744	0.441	756.87	892.64	0.744	0.442
	2	0.614	0.546	928.06	835.95	0.615	0.511

Table 13
Best RGs networks.

Networks	Rank	RGs involved
AHP	Best	M17A,KI9,M20,M18,M16,KI8,M15
SOM1 or HC1	Second	KI8,KI9,M14,M15,M18,M17A,M20
HM	Third	M1,M17A,M22,M15,M18,M17,M16

It is revealed from the Table 11 that the stations M17A, M18, M15 are common in 3 top networks whereas KI8, KI9, M20 are common to AHP and SOM1/HC1 network and M16 is common to HM and AHP network. Again AHP and SOM network has six and AHP and HM have 4 RGs common between them. Addition of M16 to the network of AHP makes it more efficient. The basic difference here between cluster methods and prioritizing methods is that cluster methods are limited to the number of sites contained in a cluster and it is difficult to find which one is most influential. But in prioritizing methods a sequence with priority is achieved. In this study both clustering methods divide the 14 RGs equally. The stations M17A, KI9, M20 are highly responsive as per storm and physiographic characteristics. The RGs selected according to AHP network establish a good rainfall–runoff model.

5.4. Flood forecasting with best RG network:

After finding the best 3 RG networks flood forecasting has been attempted at Kantamal gauge and discharge site using Fuzzy logic and ANN models. Taking these networks the inputs are fixed as discharge at Kantamal (Q_t) as per Eqs. (14)–(16).

$$Q_{t(AHP)} = f\{M17A_{(t-1)}, M17A_{(t)}, KI9_{(t-3)}, KI9_{(t-2)}, M20_{(t-1)}, M20_{(t)}, M18_{(t-2)}, M18_{(t-1)}, M18_{(t)}, KI8_{(t-3)}, M15_{(t-1)}, M15_{(t)}, M16_{(t)}, Evaporation_{(t)}\} \quad (14)$$

$$Q_{t(SOM1)} = f\{KI8_{(t-3)}, KI9_{(t-3)}, M14_{(t-1)}, M14_{(t)}, M15_{(t-1)}, M15_{(t)}, M18_{(t-2)}, M18_{(t-1)}, M18_{(t)}, M17A_{(t-1)}, M17_{(t)}, M20_{(t-1)} M20_{(t)}, Evaporation_{(t)}\} \quad (15)$$

Table 14
Performance measures of models for 1-day lead period flood forecasting.

Network	Model	Correlations		RMSE(m^3/s)		Efficiency	
		Calibration	Validation	Calibration	Validation	Calibration	Validation
AHP	ANN	0.9504	0.8775	465.132	590.366	0.9032	0.7595
	Fuzzy	0.9195	0.9105	587.67	500.212	0.8455	0.8274
SOM1	ANN	0.9478	0.8587	476.677	620.58	0.8984	0.7343
	Fuzzy	0.9116	0.8984	614.776	530.009	0.8309	0.8062
HM	ANN	0.9221	0.8471	578.372	649.684	0.8503	0.709
	Fuzzy	0.8739	0.8582	726.781	625.808	0.7637	0.73

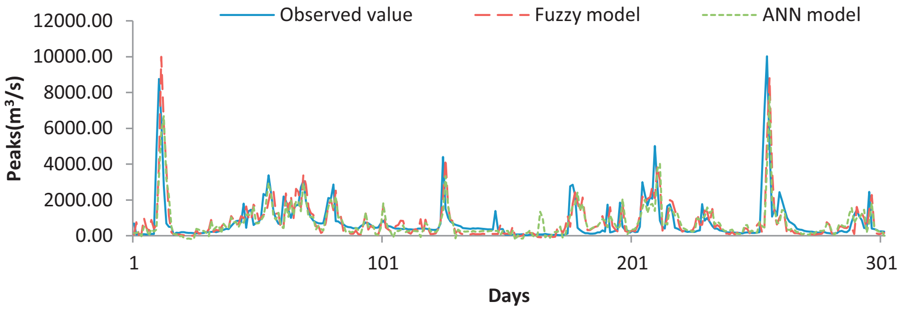

Fig. 10. Flood forecasting at Kantamal with 1-day lead time.

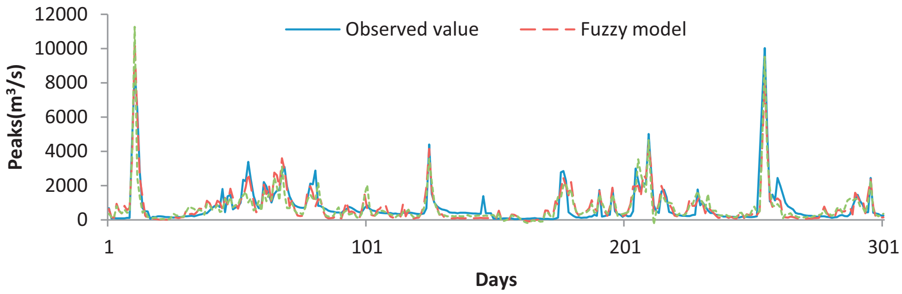

Fig. 11. Flood forecasting at Kantamal with 1-day lead time (SOM).

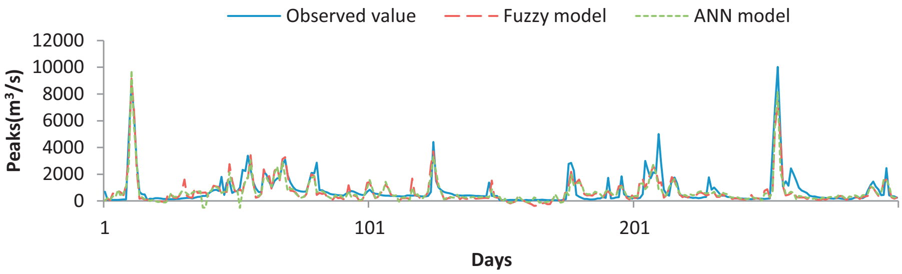

Fig. 12. Flood forecasting at Kantamal with 1-day lead time (HM).

$$Q_{t(HM)} = f\{M1_{(t-1)}, M17A_{(t)}, M22_{(t-1)}, M15_{(t-1)}, M18_{(t-2)}, M18_{(t-1)}, M18_{(t)}, M17_{(t-1)}, M16_{(t)}, \text{Evaporation}_{(t)}\}. \quad (16)$$

The inputs are put into a 3 layered feed forward network with 5 neurons in hidden and one output neuron. The attempt is made for flood forecasting at one and two day lead periods. In the same way the Takagi–Sugeno fuzzy model is also applied for the same inputs with cluster radius of 0.5. The results of the both models are shown in Table 14. However, the one day lead period results are shown here, because the two day lead period results in poor efficiency and are eliminated from the study. The Figs. 10–12 shows the results of ANN and Fuzzy models for the observed floods. It is again verified that a Fuzzy logic based flood forecasting model gives higher efficiency than ANN model.

6. Conclusions

In the flood prone basins selection of suitable RG network is a very important task for making a reliable flood forecast during the time of the failure of some of the RG stations or non receipt of data. In this paper a procedure for the design of the key RG network for flood forecasting is discussed and demonstrated through a case study using the data of a sub-basin of Mahanadi basin and the following conclusions are drawn:

(1) The study will be of use for flood forecasting during the time of distress when there will be difficulty in receiving the information. (2) Both the prioritization method and clustering methods are applied here to form the key and alternate RG network. Prioritization of RGs as per Hall method is based on the storm data, whereas AHP, HC, SOM gives importance to both physical and hydro-meteorological characteristics but AHP decides as per the merits of each characteristic. (3) Three best RG networks are established on priority basis and the information from these RGs are vital during flood forecasting. (4) Flood forecasting is possible to a reasonable efficiency in this study using atleast seven defined RGs instead of 14 stations established. (5) The AHP method supersedes other methods for finding the best RG network. (6) The forecaster can either use any of the networks subject to availability of that information with the designated efficiency of that network. (7) The Fuzzy logic developed method is best over ANN and NAM model in modeling as well as flood forecasting. (8) As the basin is a medium sized one flood forecasting for 1-day R–R lead period is feasible. (9) The results of clustering methods show that the RG's are divided into two equal sized groups. In that context prioritization methods are always better as it suggests the sequence of the network rather than dividing them into two groups. (10) The study can also be trialed with more variables, RGs and longer time series and tested in other basins.

References

Acreman, M.C., Sinclair, C.D., 1986. "Classification of drainage basins according to their physical characteristics: an application for flood frequency analysis in Scotland". J. Hydrol. Amsterdam 84, 365–380.

Agarwal, A., Lohani, A.K., Singh, R.D., Kasiviswanathan, K.S., 2013. Radial basis artificial neural network models and comparative performance. J. Indian Water Resour. Soc. 33 (4), 1–8.

Anane, M., Hammadi, K., Jellali, S., 2008. Ranking suitable sites for soil aquifer treatment in Jerba Island (Tunisia) using remote sensing, GIS and AHP-multicriteria decision analysis. Int. J. Water 4 (1/2), 121–135.

Berger, K.P., Entekhabi, D., 2001. Basin hydrologic response relations to distributed physiographic descriptors and climate. J. Hydrol. 247, 169–182.

Campolo, M., Andruessi, P., Soldati, A., 1999. River flood forecasting with a neural network model. Water Resour. Res. 35 (4), 1191–1197.

Campolo, M., Soldati, A., Andreeussi, P., 2003. Artificial neural network approach to flood forecasting in the River Arno. Hydrol. Sci. J. 35 (4), 1191–1197.

Casper, M., Gemmar, P., Gronz, O., Johst, M., Stuber, M., 2007. Fuzzy logic-based rainfall-runoff modeling using soil moisture measurements to represent system state. Hydrol. Sci. J. 52 (3), 478–490.

Cheng, K.-S., Lin, Y.-C., Liou, J.-J., 2008. Rain-gauge network evaluation and augmentation using geostatistics. Hydrol. Processes 22 (14), 2554–2564.

Chiu, S., 1994. Fuzzy model identification based on cluster estimation. J. Intell. Fuzzy Syst. 2, 267–278.

Coulibaly, P., Anctil, F., Bobee, B., 2000. Daily reservoir inflow forecasting using artificial neural network with stopped training approach. J. Hydrol. 230 (3–4), 244–257.

Dawson, C.W., Wilby, R.L., 2001. Hydrological modeling using ANN. Prog. Phys. Geogr. 25 (1), 80–208.

Dharmasena, G.T. 1997. Application of mathematical models for flood forecasting in Srilanka Destructive Water: Water-caused natural disasters, their abatement and control (proceeding of the conference held at Anaheim, California, June, 1996), *IAHS*, Publ.No. 239, 225-235.

Everitt, B., 1980. Cluster Analysis. Halsted press, a Division of John Wiley and Sons, Inc., New York.

Freeze, R.A., Cherry, J.A., 1979. Groundwater Hydrology. Prentice-Hall, pp. 604.

Fovell, R.G., Fovell, M.Y.C., 1993. Climate zones of the conterminous United States using cluster analysis defined using cluster analysis. J. Clim. 6, 165–168.

Hall, A.J., 1972. "Methods of Selection of Areal Rainfall Stations and the Calculation of Areal Rainfall for Flood Forecasting Purposes", 27. Commonwealth Bureau of Meteorology, Melbourne, pp. 10.

Harun, S., Nor, N.I.A., Kassim, A.H.M., 2001. Rainfall–runoff modelling using artificial neural network. Malays. J. Civil Eng. 13 (1), 37–50.

Hundecha, Y., Bardossy, A., Theisen, H., 2001. Development of a fuzzy logic-based rainfall–runoff model. Hydrol. Sci. J. 46 (3), 363–376.

Imrie, C.E., Durucan, S., Korre, A., 2000. River flow prediction using artificial neural network: generalization beyond the calibration range. J. Hydrol. 233 (1–4), 138–153.

Jang, J.-S.R., Sun, C.T., Mizutani, E., 2002. Neuro-fuzzy and self computing. Prentice Hall, New Delhi.

Jingyi, Z., Hall, M.J., 2004. Regional flood frequency analysis for the Gan-Ming river basin in China. J. Hydrol. 296, 98–117.

Kar, A.K., Winn, L.L., Lohani, A.K., Goel, N.K., 2011. Soft computing–based workable flood forecasting model for Ayeyarwady River Basin of Myanmar. J. Hydrol. Eng. 17 (7), 807–822.

Kar, A., Lohani, A.K., 2010. Development of flood forecasting system using statistical and ann techniques in the downstream catchment of Mahanadi basin, India. J. Water Resour. Prot. 2, http://dx.doi.org/10.4236/jwarp.2010.210105, No. 10, pp. 880-887.

Kagan, R.L., 1966. On the evaluation of representativeness of RG data G1. Geofiz. Obs. 191, 22–34.

Kalteh, A.M., 2008. Rainfall–runoff modeling using artificial neural networks (ANNs): modeling and understanding. Caspian J. Environ. Sci. 6 (1), 53–58.

Kevin, D.Y., Kibler, D.F., Benham, B.L., Loganathan, G.V., 2009. Application of analytical hierarchical process for improved selection of storm water BMPs ASCE. J. Water Resour. Plann. Manag. 135 (4), 264–275.

Kiang, M.Y., Kulkarni, U.R., Goul, M.R., Chi, R.T., Terban, E., Philppakkis, A., 1997. Improving the effectiveness of self organization map networks using a circular Kohonen layer. 30th Hawaii International Conference on System Sciences (HICSS), Volume 5. Advanced Technology Track, pp. 521.

Kohonen, T., 1982. Self-organized formation of topologically correct feature maps. Biol. Cybern. 43 (1), 59–69, http://dx.doi.org/10.1007/bf00337288.

Lekkas, D.F., Onof, C., Lee, M.J., Baltas, E.A., 2004. Application of neural networks for flood forecasting. Global NEST: Int. J. 6 (3), 205–211.

Lim, Y.H., Lye, L.M., 2003. Regional flood estimation for ungauged basins in Sarawak, Malaysia. Hydrol. Sci. J. 48 (1), 79–94.

Lin, C., 2006. Scenario deployment of the analytic hierarchy process for the radwaste repository site selection in Taiwan. In: IEEE International Conference on Systems, Man, and Cybernetics, October 8–11, 2006, Taipei, Taiwan.

Lohani, A.K., Arora, M., 1992. Design of key network stations for the purpose of flood forecasting- a case study. Hydrol. J. IAH XV(1&2), 35–44.

Lohani, A.K., Arora, M., 1995. Comparison of various precipitation network design methods in a typical basin. Hydrol. J. IAH XVIII(3&4), 99–111.

Lohani, A.K., Goel, N.K., Bhatia, K.K.S., 2005a. Real time flood forecasting using fuzzy logic. In: Perumal, M. (Ed.), Hydrological Perspectives for Sustainable Development, 1. Allied Publishers Pvt. Ltd., New Delhi, pp. 168–176.

Lohani, A.K., Goel, N.K., Bhatia, K.K.S., 2005b. Development of fuzzy logic based real time flood forecasting system for river Narmada in Central India. In: In International Conference on Innovation Advances and Implementation of Flood Forecasting Technology, ACTIF/Floodman/Flood Relief, October, 2005. Tromso, Norway www.Actif.cc.net/ conference 2005 /proceedings.

Lohani, A.K., Goel, N.K., Bhatia, K.K.S., 2006. Takagi–Sugeno fuzzy inference system for modeling stage-discharge relationshi. J. Hydrol. 331, 146–160.
Lohani, A.K., Goel, N.K., Bhatia, K.K.S., 2007a. Deriving stage—discharge—sediment concentration relationships using fuzzy logic. J. Hydrol. Sci. 52 (4), 793–807.
Lohani, A.K., Goel, N.K., Bhatia, K.K.S., 2007b. Reply to comments provided by Z. Sen on Takagi–Sugeno fuzzy system for modeling stage-discharge relationship by A. K. Lohani, N. K. Goel and K. K. S. Bhatia. J. Hydrol. 337 (1–2), 244–247.
Lohani, A.K., Goel, N.K., Bhatia, K.K.S., 2011. Comparative study of neural network, fuzzy logic and linear transfer function techniques in daily rainfall-runoff modeling under different input domains. Hydrol. Processes 25, 175–193.
Lohani, A.K., Kumar, R., Singh, R.D., 2012. Hydrological time series modeling: a comparison between adaptive neuro-fuzzy, neural network and autoregressive techniques. J. Hydrol. 442, 23–35.
Lohani, A.K., Goel, N.K., Bhatia, K.K.S., 2014. Improving real time flood forecasting using fuzzy inference system. J. Hydrol. 509, 25–41.
Lohani, A.K., Krishan, G., 2015. Application of artificial neural network for groundwater level simulation in Amritsar and Gurdaspur Districts of Punjab, India. Clim. J Earth Sci. Change 6, 4, http://dx.doi.org/10.4172/2157-7617.1000274.
Minns, A.W., Hall, M.J., 1996. Artificial neural networks as rainfall–runoff models. Hydrol. Sci. J. 41 (3), 399–417.
Modarres, R., 2009. Multi-criteria validation of artificial neural network rainfall–runoff modeling. Hydrol. Earth Syst. Sci. 13, 411–421.
Morin, G., Fortin, J.P., Sochanska, W., Lardeau, J.P., 1979. Use of principal component analysis to identify homogeneous stations for optimal interpolation. Water Resour. Res. 15 (6), 1841–1850.
Muhamad, J.R., Hassan, J.N., 2005. Khabur River flow using artificial neural networks. Al Rafidain Eng. 13 (2), 33–42.
Mukerjee, A., Chatterji, C., Singh, N., Raghuvanshi, N.S., 2009. Flood forecasting using ANN, Neuro-Fuzzy and Neuro-GA models. J. Hydrol. Eng. 14 (6), 647–652.
Nayak, P.C., Sudheer, K.P., Ramashastri, K.S., 2005. Fuzzy computing based rainfall–runoff model for real time flood forecasting. Hydrol. Processes 19, 955–968.
Patra, K.C., 2002. Hydrology and water resources engineering. Narosa Publishing House, New Delhi.
Pitlick, J., 1994. Relation between peak flows, precipitation, and physiography for five mountainous regions in the western USA. J. Hydrol. 158, 219–240.
Ponce, V., Lohani, A., Huston, P., 1997. Surface Albedo and water resources: hydroclimatological impact of human activities. J. Hydrol. Eng. 2 (4), 197–203.
Rabuffetti, D., Barbero, S., 2005. Operational hydro-meteorological working and real time flood forecasting: the Piemonte region case study. Hydrol. Earth Syst. Sci. 9 (4), 457–466.
Rajurkar, M.P., Kothyari, U.C., Chaube, U.C., 2002. Artificial neural networks for daily rainfall–runoff modeling. J. Hydrol. Sci. 47 (6), 865–877.
Rai, S.P., Sharma, N., Lohani, A.K., 2014. Risk assessment for transboundary rivers using fuzzy synthetic evaluation technique. J. Hydrol. 519, 1551–1559.
Riad, S., Mania, J., Bouchaou, L., Najjar, Y., 2004. Rainfall runoff model using artificial neural network approach. J. Math. Comput. Model. 40 (7-8), 839–846.
Saaty, T.L., 1980. The Analytic Hierachy Process. McGraw Hill, New York, pp. 20–25.
Saaty, T.L., 2008. Decision making with analytic hierarchy process. Int. J. Serv. Sci. 1 (1), 83–98.
Sinha, R., Bapalu, G.V., Singh, L.K., Rath, B., 2008. Flood risk analysis in the Kosi river basin, north Bihar using multiparametric approach of analytical hierarchy process (AHP). Indian J. Remote Sens. 36, 293-307.
Sreedharan, K.E., James, E.J., 1983. Design of rain gage network for the Chaliyar River basin. Journal of Institution of Engineers (India) 64 (CI 3), 170–174.
Sudheer, K.P., Gosain, A.K., Ramashastri, K.S., 2002. A data driven algorithm for constructing artificial neural networks rainfall–runoff model. Hydrol. Processes 16, 1325–1330.
Takagi, T., Sugeno, M., 1985. Fuzzy identification of systems and its application to modeling and control. Trans. Syst. Man Cybern. IEEE 15 (1), 116–132.
Tingsanchali, T., Gautam, M.R., 2000. Application of tank, NAM, ARMA and neural network models to flood forecasting. Hydrol. Processes 14, 243–2487.
Tokar, A.S., Markus, M., 2000. Precipitation-runoff modeling using artificial neural networks and conceptual models. J. Hydrol. Eng. 5 (2), 156–161.
Triantaphyllou, E., Mann, S.H., 1995. Using the analytic hierarchy process for decision making in engineering applications: some challenges. Int. J. Ind. Eng.: Appl. Pract. 2 (1), 35–44.
Vogel, R.M., Kroll, C.N., 1996. Estimation of base flow recession constants. Water Resour. Manag. 10, 303–320.
Yu, P.S., Chen, S.-T., 2005. Updating real time flood forecasting using a fuzzy rule based model. Hydrol. Sci. J. 50 (2), 265–278.
Valenca, I., Ludermir, T. 2009. Hybrid systems for river flood forecasting using MLP, SOM and Fuzzy system ICANN, Part-I, LNCS 5768, 557-566.
Zhang, B., Govindaraju, R.S., 2000. Prediction of watershed runoff using Bayesian concepts and modular neural networks. Water Resour. Res. 36 (3), 753–762.
Zadeh, L.A., 1965. Fuzzy Sets. Inf. Control 8, 338–353.