COMPLETION REPORT

(APPENDIX-5)

TITLE OF THE PROJECT

ARTIFICIAL NEURAL NETWORKS FOR WATER RESOURCES PLANNING: AN INNOVATIVE APPROACH

By

Dr K. Srinivasa Raju
Associate Professor
Civil Engineering Group

Submitted to

Indian National Committee on Hydrology
Roorkee

BIRLA INSTITUTE OF TECHNOLOGY& SCIENCE, PILANI
HYDERABAD CAMPUS
March , 2010

Appendix 5 Format for Completion Report

1. NAME AND ADDRESS OF THE INSTITUTE

Name	: Birla Institute of Technology and Science-Pilani, Hyderabad campus
Address	: Jawahar Nagar Village, Shamirpet Mandal, Ranga Reddy District, Andhra Pradesh, India
Telephone	: 040-6630 3999 (General Administration)
Fax	: ---
Web	: http://www.bits-hyderabad.ac.in/

2. NAME AND ADDRESS OF THE PI AND OTHER INVESTIGATORS

Name	: Dr Komaragiri Srinivasa Raju
Designation	: Associate Professor
Address	: Civil Engineering Group Birla Institute of Technology and Science-Pilani, Hyderabad campus Jawahar Nagar Village, Shamirpet Mandal, Ranga Reddy District, Andhra Pradesh, India
Telephone	: 040- 6630 3519 (Office)
Fax	: ---
Email	: ksraju@bits-hyderabad.ac.in

3. TITLE OF THE SCHEME

Artificial Neural Networks for Water Resources Planning: An Innovative Approach
(Ref: Project No. 23/51/2006-R&D/203-14 dated January 23, 2006)

4. FINANCIAL DETAILS

Sanctioned Cost: Rs. 4,62,000/-

Amount Released Rs. 2,94,000 + Rs.72,000/-

(Interest has also accumulated from time to time on the amount released)

Expenditure: Rs. 2,98,426/-

Unspent balance Rs. 3,312/- + Rs. 72,000 = Rs. 75,312

5. ORIGINAL OBJECTIVES AND METHODOLOGY AS IN THE SANCTIONED PROPOSAL

1. Data collection of various stream flow and watershed parameters includes climatic, topographic, land use etc.

2. Classification of chosen river basin into various homogeneous regions with the above collected data using Artificial Neural Network classification techniques, namely, (a) Adaptive Resonance Theory (ART) and (b) Kohenon Neural Network (KNN).
3. Develop a regional relationship between stream flows (runoff) and watershed variables using Artificial Neural Networks techniques namely, (a) Feed forward network using back propagation (b) Radial basis function network.
4. Suitability of Artificial Neural Network algorithms, namely, (a) Adaptive Resonance Theory (ART) and (b) Kohenon Neural Network (KNN) (c) Feed forward network using back propagation (d) Radial basis function network for the chosen river basin and comparison of results with existing methods such as regression (for prediction), cluster analysis (for classification).
5. Identify various pitfalls in ANN methodology and develop methods to overcome the same for further improvements.
6. Application of above methodology to Luni river basin in Rajasthan

6. ANY CHANGE IN THE OBJECTIVES DURING THE OPERATION OF THE SCHEME

There is no change in the objectives. However, some more techniques were added/ replaced based on the suggestions during R & D sessions at Varanasi and Udaipur.

Fuzzy Cluster Analysis is explored in the present study as per suggestion by experts at R & D session due to its advantage of requiring less number of input parameters and its ability of assigning every microwatershed with partial membership in each group in place of Adaptive Resonance Theory (ART). Accordingly Fuzzy Cluster Analysis is used.

Similarly, Suitability of cluster validation indices, namely, Davies-Bouldin, Dunn's and Silhouette to determine the optimum number of clusters/groups of microwatershed is explored.

7. DATA COLLECTED AND USED IN THE ANALYSIS WITH SOURCES

Case Study: Kherthal watershed, Bali Panchayat Samithi, Pali District, Rajasthan

Data: Various stream flow and watershed parameters such as catchment area, basin length, drainage density, channel slope, landuse, mean basin elevation, rainfall and discharge are measured/estimated/inferred from various data sources such as State Remote Sensing Application Center (SRSAC), Jodhpur. In addition Bifurcation ratio (R_b), Stream frequency (F_u), Length of overland flow (L_o), Form factor (R_f), Shape factor (B_s), Elongation ratio (R_e), Circulatory ratio (R_c), Compactness coefficient (C_c), Texture ratio (T) which are mainly morphological parameters are estimated.

Sources of Data: IRS-LISS-III imageries of the case study area; Survey Of India (SOI) toposheets 45 H/1, 45 H/5 of the scale 1:50000; Ground Water Atlas of Rajasthan 2000; Watershed Atlas of Rajasthan 2000; These were analysed with the help of State Remote Sensing Application Center (SRSAC), Jodhpur.

8. METHODOLOGY FOLLOWED (OBSERVATIONS, ANALYSIS, RESULTS AND INFERENCES)

Watershed can be considered as a unit/block for planning of water and soil conservation programmes for improving the water resources in the region. This also enables to plan sustainable development in the region. Significant factors that affect planning and development of watershed are physiography, drainage, morphology, soil, land use/land cover and available water resources. This complexity necessitates establishment of relationships between different parameters thereby producing concise representation of the system's behavior in the perspective of water resources system. Systematic delineation/ classification of microwatersheds also facilitates the choice of strategies appropriate for each group in the region in respect of agricultural practices. This paves the way for efficient utilization of resources (State Water Policy, Government of Rajasthan 1999). In this context, Artificial Neural Networks (ANN) and other related/ relevant methodologies are gaining importance as a classification and prediction solution methodologies to analyze situations discussed above. These methodologies are also helpful for drainage delineation, morphometric analysis, watershed development and management and prioritization of microwatersheds.

With this background, objectives of the study are formulated as follows:

1. Data collection: Various stream flow and watershed parameters such as catchment area, basin length, drainage density, channel slope, land use, mean basin elevation, rainfall and discharge are measured/estimated/inferred from various data sources. In addition Bifurcation ratio (R_b), Stream frequency (F_u), Length of overland flow (L_o), Form factor (R_f), Shape factor (B_s), Elongation ratio (R_e), Circulatory ratio (R_c), Compactness coefficient (C_c), Texture ratio (T) which are mainly morphological parameters are estimated.
2. Classification of chosen watershed into various homogeneous regions using Kohonen Neural Networks (KNN), Fuzzy Cluster Analysis (FCA), Cluster Analysis (CA) based on morphological and other parameters. Fuzzy Cluster Analysis is explored in the present study as per suggestion of experts during R & D session, due to its advantage of requiring less number of input parameters and its ability of assigning every

microwatershed with partial membership in each group. In the present study, Cluster Analysis is used as the bench marking technique to compare the outcome of above two techniques.

3. Exploring the suitabilitiy of cluster validation indices, namely, Davies-Bouldin, Dunn's and Silhouette to determine the optimum number of clusters/groups of microwatersheds.
4. Development of relationship between discharge and morphological and other related parameters (as per suggestion during R & D session) using Artificial Neural Networks techniques namely, Feed Forward with Back Propagation (FFBP) and Radial Basis Function (RBF). Multiple Linear Regression is used as the bench marking technique.
5. Sensitivity analysis of selected parameters in some of the employed techniques.

The case study considered for the application of the above methodologies is Kherthal watershed, Bali Panchayat Samithi, Pali district, Rajasthan, India. Detailed literature review is presented in annexure-1.

Following points are evident from the literature review (presented in annexure-1) on studies of watershed development and management at microlevel.

1. No studies are reported in classification aspects using morphological parameters.
2. Cluster validation indices are not explored.
3. No studies are reported in prediction aspects using morphological parameters.

Adequately addressing the above limitations is the basis for formulating the objectives for the present study.

TECHNIQUES EMPLOYED

CLASSIFICATION METHODOLOGY

Rao and Srinivas (2008) discussed important issues related to clustering such as choice of clustering algorithm, choice of appropriate attributes for clustering, selection of an objective function, choice of dissimilarity (or distance) measure, appropriate initialization of the clustering algorithm and selection of appropriate number of clusters in the data.

In the present study, practical applicability and suitability of three classification techniques, namely, Kohonen Neural Networks (KNN), Fuzzy Cluster Analysis (FCA), and Cluster Analysis (CA) are explored for grouping 25 microwatersheds which are part of Kherthal watershed, Bali Panchayat Samithi, Pali District, Rajasthan. These techniques are explained in brief below.

Kohonen Neural Networks (KNN)

Kohonen Neural Networks (KNN) is a self organizing mapping technique with two layers, input and output. Each layer is made up of neurons. These are based on unsupervised classification and consist of competitive layers that use learning rule to group inputs (Rao and Srinivas 2008). The neurons of the competitive layer learn to recognize groups of similar input vectors. The number of neurons in input layer, M, is equal to the dimensionality of the input vectors and the number of neurons in the output layer, N, is determined by the number of groups into which the input data will be partitioned. Each neuron in the output is interconnected with all those in the input layer by a set of weights or a weight vector, e.g., the j^{th} output neuron has a weight vector connecting to input neurons, $w_j = \{w_{ji}\}, i = 1,2,....,M$. The function of an input neuron is to transmit input data to the next layer, whereas an output neuron calculates the Euclidean distance between its weight vector w_j and input vector X' to measure their similarity. The main objective of the Kohonen networks is to transform an incoming vector with arbitrary dimension into a one or two dimensional discrete map, and to perform this transformation adaptively in a topologically ordered fashion (Kohonen 1989; Liong et al. 2004). Important input parameters for KNN are learning rate, conscience rate, and number of epochs and tolerance criterion. Figure 1 presents schematic diagram of Kohonen Neural Networks relevant to the planning problem where M and N are 10 and 5 respectively.

Fuzzy Cluster Analysis (FCA)

Fuzzy Cluster Analysis is a clustering technique in fuzzy environment wherein each data point belongs to a cluster to some degree, specified by a membership grade. The algorithm is based on minimizing an objective function that represents the distance from any given data point to a cluster center, weighted by that data point's membership grade. In other words, the objective of the technique is to represent the similarity a point shares with each cluster with a membership function, whose value lies between zero and one. Each sample has a membership in every cluster (Ross 1995) but degree of membership varies from cluster to cluster (between zero to one). The sum of the membership values for each dataset will be equal to 1. A brief methodology of FCA is given below:

1. Normalize the data
2. Choose the number of clusters c ($2 \leq c \leq n$, where n is number of data points), number of iterations and termination criteria
3. Formulate initial fuzzy partition matrix
4. Compute cluster centers for each iteration and ascertain the Euclidean distance between the dataset and the cluster centers.
5. Update the fuzzy partition matrix for each iteration.

If fuzzy partition matrix between two successive iterations is less than the specified termination criteria, the algorithm stops. Otherwise steps 2 to 5 are to be repeated until the above requirement is satisfied. Important input parameters for FCA are number of iterations and tolerance criterion. More detailed description of FCA is available in Ross (1995), Jingyi and Hall (2004), and Rao and Srinivas (2006a), Rao and Srinivas (2008).

Cluster Analysis (CA)

Cluster Analysis partitions datasets into relatively homogeneous groups. K-means clustering algorithm (Jain and Dubes 1988) is used to minimize within cluster sums of squares of differences to obtain the final partitions. In K-means technique each cluster is represented by its mean of feature vectors within the cluster (Rao and Srinivas 2008).

In this technique, datasets are grouped so that each dataset is assigned to one of the fixed number K of groups. The sum of the squared differences of each criterion from its assigned cluster mean is used as the objective function. Datasets are transferred from one cluster to another, so that, within- cluster, the sum of squared differences decreases. In a pass through the entire dataset, if no transfer occurs, the algorithm stops. The total squared error value E_K for cluster group K, is given by

$$E_K = \sum_{k=1}^{K} e_k^2 \qquad (1)$$

where e_k = Error value for each cluster group k.

Important input parameters for CA are number of epochs and tolerance criterion.

Silhouette Index

The Silhouette index (Rousseeuw 1987) computes the silhouette width for each sample, average silhouette width for each cluster and overall average silhouette width for the total data set. Each cluster is represented by a Silhouette, based on the comparison of its tightness and separation. The average silhouette width is used to decide how good the number of selected clusters is. Silhouette width *S(i)* is expressed as:

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (4)$$

where *a(i)* –average distance (Linear Euclidean) from i^{th} dataset to all other datasets in the group k ; *b(i)* – minimum average distance (Linear Euclidean) from i^{th} dataset to all the datasets in another group j (j=1,2,..K; j $\neq$ k). $S(i)$ varies between -1 to 1. Three types of inferences are possible: If S(i) is close to 1, i^{th} dataset has been assigned to an appropriate cluster; If S(i) is close to –1, i^{th} dataset has been misclassified; If S(i) is approximately zero, the i^{th} dataset lies at equal distant from the two clusters. For the given K clusters, the overall average silhouette width is the average of the silhouette widths for all the datasets in the group. The partition with the maximum overall average silhouette width is taken as the optimal partition or indicative of the optimum number of clusters.

PREDICTION METHODOLOGY

In the present study, practical applicability and suitability of three prediction techniques, namely, Feed Forward with Back Propagation, Radial Basis Function and Multiple Linear Regression are explored for developing relationship between discharge and morphological and other related parameters for Kherthal watershed, Bali Panchayat Samithi, Pali District, Rajasthan. These techniques are explained in brief below.

Feed Forward Neural Networks

Feed Forward with Back Propagation (FFBP) algorithm of supervised classification is a technique to train feed forward models and corresponding network consists of layers of parallel processing elements i.e., neurons. The algorithmic process in FFBP consists of fixing the network architecture, training and learning rules and testing the network. Learning rules specify an initial set of assumed weights and indicate how the weights should be adapted to improve performance. The activation functions (sigmoid, tan sigmoid or log sigmoid) are useful to transform input signal into an output signal.

In Feed forward models, the outputs can be sent only to the immediate next layers. Each layer being fully connected to the preceding layer by interconnection strengths/weights (Rumelhart et al. 1986; Kisi 2008). Figure 2 illustrates a three-layer FFBP consisting of input layer (10 neurons), hidden layer (6 neurons) and output layer (1 neuron), which is used in the present planning problem. Assumed initial weight values are updated epoch wise (iteration) during training process in which the predicted outputs are compared with the known/ observed outputs, and errors if any, are back propagated. The total error based on the squared difference between predicted and actual outputs for data set is computed. The process will be terminated after reaching error tolerance/ number of epochs whichever is earlier. This process is necessary to determine the final weights.

Important parameters that are required for this network are architecture, learning rate, type of activation function, definition of error, number of epochs etc (Nagesh Kumar 2004). Detailed procedures for training and testing the network (including the mathematical expressions) are available in ASCE (2000a).

Radial Basis Function

Radial Basis Function (RBF) network with feed-forward architecture consists of input layer, hidden layer and output layer (total 3 layers) with number of neurons in each layer similar to FFBP. Radial Basis Function is based on the self-organized characteristics whereas FFBP is based on supervised classification. RBF methodology allows for adaptive determination of the hidden neurons during training of the network. The hidden layer performs a fixed nonlinear transformation with no adjustable parameters (Zhang and Kushwaha 1999; Sharma et al. 2003).

Hidden layer in RBF consists of a number of nodes and a parameter vector called center i.e., weight vector of the hidden layer. For each node, the Euclidean distance between the center and the input vector of the network is computed and transformed by a nonlinear function that determines the output of the nodes in the hidden layer. The output layer then combines these results in a linear fashion. The common nonlinearity used in Radial Basis Function is Gaussian Kernel function where response is unity if this distance is zero, and decays to zero when the distance is greater than the spread. The response of j th hidden neuron h_j to an input signal X is given by

$$h_j = f\left[-\frac{\|X-\mu_j\|}{2\sigma_j^2}\right], \quad j=1,2,\ldots\ldots J \qquad (5)$$

where $\|.\|$ is the Euclidean distance, μ_j and σ_j are center and spread of hidden neuron j respectively. The responses are multiplied by the interconnection weights W, between the hidden and output layers. Each unit in the output layer then makes a linear transformation on the data from the hidden layer. The response y_k of neuron k can be described as:

$$y_k = \sum_{j=1}^{J} h_j W_{kj} \qquad ; k = 1,2,\ldots K \qquad (6)$$

where W_{kj} is the connection weight between the hidden neuron j and output neuron k. Important parameters required for this network are architecture, algorithm, epochs, tolerance criterion and spread constant. More details about Radial Basis Function are available in ASCE (2000a), Kartalopoulos (2002), Sudheer and Jain (2003) and Sharma et al. (2003).

Multiple Linear Regression

Regression is the statistical technique used to investigate the relationship between a dependant variable y and one or more independent variables x. There are many types of regression techniques that can be explored (www.mathworks.com). A linear regression model that formulates the relationship between the dependant variable and more than one independent variable is called a multiple regression model. Mathematical expression for multiple linear regression is

$$y = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_n x_n + e \qquad (7)$$

where $a_1, a_2, \ldots a_n$ are coefficients of $x_1, x_2, \ldots x_n$; a_0 is the constant and e is the error.

<u>CASE STUDY AREA</u>

Kherthal watershed, Bali Panchayat Samithi, Pali District, Rajasthan is considered as a case study and is located between 24°51' to 24°58' North latitudes and 73°8' to 73°19' East longitudes. Area of watershed is 158.93 km^2 (Ground Water Atlas of Rajasthan 2000; Watershed Atlas of Rajasthan 2000). Some of the data (mentioned in objective 1) are inferred from IRS-LISS-III imageries of the case study. In addition, Survey Of India (SOI) toposheets 45 H/1, 45 H/5 of the scale 1:50000 are also used. Figure 3 presents location map of the Kherthal watershed. Figure 4 presents detailed map of Kherthal watershed showing the 25 microwatersheds.

The area is in semi-arid zone. The climate of the watershed is dry. It is very hot during summer and cold during winter. Maximum temperature is 45°C and minimum is 1°C. January is the coldest while May and June are the hottest months. Normal rainfall in the region is 490 mm. Average number of rainy days in a year is only 22 (Ground Water Atlas of Rajasthan 2000). Crops grown in Rabi season are Wheat and Mustard whereas Bajra, Pulses, Moong are grown in Kharif season. Ground water is the major source of irrigation and groundwater potential in the region is 40 to 80 m^3/day. There are 15 to 18 wells maintained by the Government. Quality of ground water is significantly influenced by semi-arid climate and hydrogeological diversity. Salinity, sodicity and fluoride are the major factors effecting the ground water quality. Quality of water varies from potable to unpotable with E_c less than 2000 $\mu S/cm$. Nitrate content ranges from 50-100 mg/l. Fluoride content ranges from 1.5 to 3 mg/l. Geologically the area consists of Granite whereas geomorphologically it consists of structural and denudational hills, pediments, buried pediments and valley fields. Figures 5 to 9 present land use/ land cover, slope, drainage, geomorphology and lithology aspects respectively (Ground Water Atlas of Rajasthan 2000; Watershed Atlas of Rajasthan 2000).

Table 1 presents microwatershedwise area, perimeter, channel slope, mean basin elevation and basin length of Kherthal watershed. Minimum and maximum values of area, perimeter, slope, mean basin elevation are respectively (0.50 km^2, 50.500 km^2), (3.077 km, 38.772 km), (1 %, 38 %), (345 m, 772 m).

MORPHOMETRIC ANALYSIS

Morphometric analysis provides quantitative description of the basin geometry (Strahler 1957; Nantiyal 1994). Relevant terminology is explained in brief below for easier understanding of the analysis (Garde 2006).

Stream order: It is based on hierarchic ranking of streams proposed by Strahler (1957). The first order streams have no tributaries whereas second order streams have only first order streams as tributaries. Similarly, third order streams have first and second order streams as tributaries.

Stream Length (L): It is the total length of all streams in the microwatershed (km).

Drainage Density (D_d): Drainage density represents the closeness of channel spacing. It is the measure of the total length of the stream segments of all orders per unit area (km^{-1}).

Bifurcation Ratio (R_b): It is the ratio of the number of streams of a given order to the number of streams of the next higher order (Schumn 1956). Horton (1945) considered bifurcation ratio as an index of relief and dissections (No unit).

Stream Frequency (F_u): It is the total number of stream segments of all orders per unit area (km^{-2}).

Length of Overland Flow (L_o): It is the length the water travels over the ground before it gets concentrated into definite stream channels. It approximately equals to half of reciprocal of drainage density (km).

Form Factor (R_f) and Shape Factor (B_s): It is ratio of basin area to square of the basin length (No unit). The value of form factor would always be less than 0.7854 for a perfectly circular basin. Shape factor is inverse of form factor.

Elongation Ratio (R_e): It is the ratio between the diameter of the circle of the same area as the drainage basin and the maximum length of the basin. A circular basin is more efficient in run-off discharge than an elongated basin (Singh and Singh 1997). The value of elongation ratio generally varies from 0.6 to 1.0 (No units) .

Circulatory Ratio (R_c): It is ratio of the area of the basin to the area of the circle having its circumference equal to the perimeter of the basin and is influenced by the length and frequency of streams, relief and slope of the basin (No units).

Compactness coefficient (C_c): It is defined as the ratio of perimeter to square root of the basin area with a multiplying factor of 0.2821 (No units).

Texture Ratio (T): It is the ratio of the number of stream segments of first order and the perimeter of that watershed (km^{-1}).

In the present study, catchment area, basin length, drainage density, channel slope, landuse, mean basin elevation, rainfall and discharge are measured/estimated/inferred from various data sources. In addition Bifurcation ratio (R_b), Stream frequency (F_u), Length of overland flow (L_o), Form factor (R_f), Shape factor (B_s), Elongation ratio (R_e), Circulatory ratio (R_c), Compactness coefficient (C_c), Texture ratio (T) which are mainly morphological parameters are estimated from the above data.

It is observed from Table 2 that total number of streams is minimum of 1 with length 1.05 km in microwatershed number 1 whereas it is maximum of 319 with length 171.31 km in case of microwatershed number 2. Mathematical expressions for the above parameters are presented in Table 3. Sample calculations in respect of microwatershed number 3 are also presented in Table 3. Ten parameters, mainly morphological, namely, Drainage density (D_d), Bifurcation ratio (R_b), Stream frequency (F_u), Length of overland flow (L_o) , Form factor (R_f), Shape factor (B_s), Elongation ratio (R_e), Circulatory ratio (R_c), Compactness coefficient (C_c) and Texture ratio (T) are considered as the classification criteria for grouping microwatersheds

(Chopra et al. 2005; Ratnam et al. 2005). These are computed using the collected data for the 25 microwatersheds and presented in Table 4. It is observed that Drainage density varies between (0.384, 7.985) among 25 microwatersheds. Similarly the lower and upper values are: Bifurcation ratio (0, 3.764), Stream frequency (0.364, 22.25), Length of overland flow (0.063, 1.303) , Form factor (0.341, 0.638), Shape factor (1.566, 2.934), Elongation ratio (0.658, 0.901), Circulatory ratio (0.185, 0.664), Compactness coefficient (1.228, 2.323), Texture ratio (0.129, 4.330). Figures 10 to 14 present variation of drainage density, form factor, shape factor, compactness coefficient and texture ratio for the 25 microwatersheds.

RESULTS AND DISCUSSION

Classification

Normalization

Estimated parameters as presented in Table 4 are normalized based on the studies by Biswas et al. (2002), Ratnam et al. (2005) and discussion with experts. Normalization/ relative comparison of criterion j (Pomerol and Romero 2000) for microwatershed i is defined as

$$y_{ij} = \frac{x_{ij}}{(x_{jideal})} \tag{8}$$

where x_{ij} is j^{th} criterion for the i^{th} microwatershed, x_{jideal} is ideal value of j^{th} criterion among the 25 microwatersheds. For example, in case of drainage density maximum value is ideal value i.e., 7.985 whereas in case of compactness coefficient minimum value is ideal i.e., desirable, accordingly it is taken as 1.228. In case of microwatershed 1, normalized value for drainage density is 0.384/7.985 resulting 0.048 whereas in case of compactness coefficient it is 1.321/ 1.228 = 1.076 (Ratnam et al. 2005). Similarly other values are normalized and presented in Table 5. Even though, Form factor, Shape factor, Bifurcation ratio, Elongation ratio, Circulatory ratio, Compactness coefficient do not have any units, their rescaling is made because of the differences in their variance, relative magnitude etc (Rao and Srinivas 2008). This also helped to maintain the consistency and uniformity in the methodology. These normalized data were used for classifying the microwatersheds.

Determination of optimal number of clusters and Application of CA

An effort was made to ascertain the optimal number of groups for the present problem. Cluster Analysis (www.mathworks.com) is used as the basis for this purpose due to its advantage of less parameter requirement and its wider applicability and acceptability to various case studies. The analysis is performed for 3 to 6 clusters/groups (total 4 in number) with 1000 iterations. The stopping criterion was set as the difference of the current objective function value from the value in the previous iteration which is to be less than the tolerance value. Equal weight is assigned to all the criteria.

Davies Bouldin index (Davies and Bouldin 1979), Dunn's index (Dunn 1974), Silhouette index (Rousseeuw 1987) were used to find the optimum number of clusters. The Cluster Validity Analysis Platform (CVAP) developed by Kaijun (2008) is used in the present study. The CVAP is based on a Graphical User Interface (GUI) for cluster validation, integrating 18 validity indices and seven clustering algorithms. Results of three validation indices computed based on Cluster Analysis as classification algorithm are presented in Table 6.

Optimum index value is marked with asterisk (*) along with corresponding number of groups. It is observed that Silhouette index prefers 6 to be the optimum number of groups, with maximum value as 0.33577. Similarly, for DB Index optimum number of groups is 6 with index value of 0.84115 whereas for Dunn's it is 4 with index value 1.6704. It can be seen that all the indices do not indicate the same number of groups as optimum, which can be expected, as each of them finds its optimum using a different algorithm. It should be understood that these indices can be used as the basis to make an informed choice about the number of groups. Accordingly, five groups were chosen for grouping the microwatersheds based on the discussion with experts and practical aspects in the field (Watershed Atlas of Rajasthan 2000). Same number of groups i.e., 5 was adopted for further analysis i.e., in utilizing in KNN and FCA.

Application of CA

Table 7 presents classification of microwatersheds obtained based on CA methodology with five as optimum groups. On adopting CA methodology, the numbers of microwatersheds falling in group numbers 1 to 5 are 3, 6, 6, 6, 4. Values in parenthesis in Table 7 represent total number of microwatersheds in that group. It is observed that 2 groups are having three and four microwatersheds each whereas other groups are having 6 each. Figure 4 presents homogeneous regions obtained by Cluster Analysis technique. Values in parenthesis are corresponding group numbers. Similarly, effort is also made to know the occupancy rate of microwatersheds in case of increase in groups i.e., 6, 7, 8. It is observed that microwatersheds have occupied all the groups even though their proportion in each group is different.

Application of KNN

Kohonen Neural Networks (www.mathworks.com) methodology is used to classify microwatersheds. The number of nodes in the input layer is equal to 10 as shown in Figure 1, i.e., Drainage density, Bifurcation ratio, Stream frequency, Length of overland flow, Form factor, Shape factor, Elongation ratio, Circulatory ratio, Compactness coefficient, Texture ratio whereas nodes in Kohonen output layer are five (optimum number of clusters). Equal weightage is assigned to all the criteria. The parameters used for training the algorithm are the number of groups five, learning rate 0.01, conscience rate 0.001, number of epochs 1000 and elapsed time 300 sec. Initially weight values (connection strength between output and input neuron) are assumed as 0.5 by the model. The algorithm terminates after reaching prespecified number of epochs or elapsed time whichever occurs earlier. Table 8 presents weight values obtained after such process. It is also observed that final weight values obtained after 1000 epochs for chosen parameters are no where matching with the initially assumed weight values of 0.5. It is observed from Table 8 that final weight values for drainage density vary from 0.2965 to 0.6012. These are (0.1886, 0.6371), (0.2152,0.5446), (0.0875,0.3266), (1.2217, 1.6642), (1.1343,1.5536), (1.1035, 1.2888), (1.4661, 3.2737), (1.0468, 1.5982), (0.1626, 0.5918) respectively for Bifurcation ratio, Stream frequency, Length of overland flow, Form factor, Shape factor, Elongation ratio, Circulatory ratio, Compactness coefficient and Texture ratio.

Table 7 presents classification of microwatersheds obtained based on KNN methodology. Out of the targeted 5 groups, all microwatersheds have fallen into 5 groups. The numbers of microwatersheds falling in these 5 groups are 4, 5, 5, 6, 5 respectively with total error value of 6.42. It may be noticed that distribution of microwatersheds amongst the five groups is almost even.

In KNN, the learning rate, for a given conscience rate and the number of epochs, plays a major role. An effort is made to observe their effect by conducting sensitivity analysis. Table 9 presents squared error values for various epochs (1000, 3000, 5000) and learning rates (0.01, 0.1 to 0.9). It is observed from Table 9 that squared error values for learning rate 0.1 to 0.9 vary (6.4085 to 10.906), (6.4069 to 8.1350), (6.3229 to 14.376) for 1000, 3000, 5000 epochs respectively. It is also observed that maximum squared error value occurs at the learning rate of 0.9 for all three variations of epochs. It is also observed that 12 microwatersheds have fallen into one group where as remaining 13 have fallen into other groups for 5000 epochs and learning rate of 0.9. Microwatersheds divided equally in each group i.e., 5, for 3000 and 5000 epochs and for learning rate of 0.01, whereas they divided into 5,4,5,5,6 in all three variations of epochs for learning rate 0.1. It is also observed that even though all five groups are occupied by microwatersheds for any of the above scenarios, most of the times, number of microwatersheds in each group vary from scenario to scenario.

Similarly, effort is also made to know the occupancy rate of microwatersheds in case of increase in groups to 6, 7, 8. It is observed that microwatersheds have occupied all the groups. However, their proportion in each group is different based on epochs. For example for 6 groups and 1000 epochs, it is (4,4,6,3,4,4) with squared error value of 5.5439; for 3000 epochs, it is (5,5,3,4,4,4) with squared error value of 5.2379; for 5000 epochs, it is 4,4,4,4,5,4 with squared error value of 5.5390. For 7 groups, it is (3,4,4,3,7,1,3) with squared error value of 5.2702 for 1000 epochs; (3,8,1,4,3,3,3) with squared error value of 5.6388 for 3000 epochs; (3,2,3,2,6,4,5) with squared error value of 5.2889 for 5000 epochs respectively. Similarly for 8 groups, it is (6,2,2,3,5,1,3,3) with squared error value of 5.992 for 1000 epochs; (3,3,4,5,3,2,3,2) with squared error value of 4.4833 for 3000 epochs; (2,1,3,4,4,3,5,3) with squared error value of 4.3507 for 5000 epochs respectively.

It is observed that effects of various learning rates for given number of epochs are significant on the squared error value even though no fixed trend is observed. Above extensive sensitivity analysis indicated that careful selection of parameters is very much important for obtaining meaningful results in KNN.

Application of FCA

Fuzzy cluster analysis (www.mathworks.com) methodology is used to classify microwatersheds. The membership value in each group indicates the probability for the microwatershed to be clustered in that specific group (Rao and Srinivas 2008). Membership values of 25 microwatersheds under each of the 5 groups (optimal number of clusters) are presented in Table 10. The group, which is having the highest membership value among the 5 groups, is the representative group for that microwatershed. For microwatershed 1, membership values for the 5 groups are 0.0857, 0.2712, 0.1608, 0.2028, 0.2795. The sum of these values should be equal to 1 (Ross 1995). The representative group for the microwatershed no. 1 is group 5 (having the maximum membership value of 0.2795). Similarly all other microwatersheds were analyzed and grouped. Numbers of microwatersheds falling in cluster groups 1 to 5 are 2, 7, 5, 5, 6 respectively. The minimum number of microwatersheds in group 1 is 2 whereas maximum is 7. It is observed that none of the 5 groups are empty. This may be due to the advantage of FCA which allows each dataset to have partial membership in all clusters. The microwatershed with the highest membership value in a group is the representative station for that group. Table 7 presents classification of microwatersheds obtained based on FCA methodology.

Similarly, effort is also made to know the occupancy rate of microwatersheds in case of increase in groups to 6, 7, 8. It is observed that microwatersheds have occupied all the groups even though their proportion in each group is different.

It is observed from Table 7 that 13 microwatersheds out of 25 are common in FCA, KNN and CA (52%); 17 microwatersheds out of 25 are common in FCA and CA (68%); 16 out of 25 are common in FCA and KNN (64%); 15 out of 25 are common in KNN and CA (60%).

Inferences: Classification Aspects

Three classification methodologies, namely, Kohonen Neural Networks, Fuzzy Cluster Analysis and Cluster Analysis were employed to group 25 microwatersheds in Kherthal watershed into homogeneous groups. Ten parameters, mainly, morphological, namely, Drainage density (D_d), Bifurcation ratio (R_b), Stream frequency (F_u), Length of overland flow (L_o), Form factor (R_f), Shape factor (B_s), Elongation ratio (R_e), Circulatory ratio (R_c), Compactness coefficient (C_c), Texture ratio (T) are used for the classification (Chopra et al. 2005; Ratnam et al. 2005). Optimal number of groups is chosen, based on Davies-Bouldin, Dunn's and Silhouette indices. The results of FCA, KNN, and CA approaches are analyzed and compared.

It is concluded that clustering Algorithms like CA, KNN, FCA prove to be vital for the efficient clustering of the datasets. Without clustering algorithms, the process of clustering would have been very cumbersome as there is a threshold beyond which the difference between the two alternatives is imperceptible to manual capabilities when compared to that of the machine. Clustering algorithms can form the groups in a small fraction of the time that is required for manual grouping, particularly if a long list of criteria is associated with data set (Jain and Dubes 1988). Upper and lower limits of number of groups for efficient working of Clustering Algorithms and validation algorithms are to be decided prori to choose the appropriate number of groups. In this regard knowledge of case study with physical characteristics, agro-climatic regions, and geographical boundaries is pertinent. Employed cluster validity indices work on two principles i.e., separation and compactness. For the clustering to be good, the compactness must be high and the separation should be as much as possible. However, it has been observed that for a number of times, the upper limit turns out to be the optimum. This happens because higher number means that the number of datasets in each group is less and so is their distance from the mean, which ensures higher compactness. As a result, the Cluster Validation Indices tend to prefer a higher number as optimum. But, this may not be always true and quite often, a number lesser than the upper limit is indicated by the cluster validation indices. Choosing more than one clustering algorithm and multiple cluster validation indices enhances the decision making ability to choose the right clustering and number of groups, as different algorithms work with different methodologies. This also increases applicability and understandability of the given situation. In the present study, optimal number suggested by the three validation indices is taken as the basis and Cluster Analysis methodology can be used as the basis for clustering microwatersheds due to its advantage of less parameter requirement as compared to two other classification techniques and its compatibility with cluster validation indices.

The present study has various new features where classification techniques and cluster validation techniques are applied to the morphological studies first time for a real case study of Kherthal watershed. The following inferences are drawn from the study:

1. Cluster Analysis methodology can be used as the basis for clustering microwatersheds due to its advantage of less parameter requirement as compared to two other classification techniques and its compatibility with cluster validation indices.
2. Methodology of determining optimal number of clusters based on Davies Bouldin, Dunn's and Silhouette is found to be useful and can be used as the basis for systematic delineation of microwatersheds.
3. It is observed that 13 microwatersheds are common in FCA, KNN and CA (52%); 17 microwatersheds are common in FCA and CA (68%); 16 in FCA and KNN (64%); 15 in KNN and CA (60%).
4. It is observed that effect of various epochs (1000, 3000, 5000) and learning rates (0.01, 0.1 to 0.9) on squared error values is significant even though no fixed trend is observed.
5. Classification methodology based on morphometric analysis can be extended for wider applications of watershed development and management.
6. Fuzzy Cluster Analysis can be explored as the basis for cluster validation techniques in further studies, along with cost analysis and homogeneity measures.

RESULTS AND DISCUSSION OF PREDICTION ASPECTS

Eleven parameters, namely, Drainage density (D_d), Bifurcation ratio (R_b), Stream frequency (F_u), Length of overland flow (L_o), Form factor (R_f), Shape factor (B_s), Elongation ratio (R_e), Circulatory ratio (R_c), Compactness coefficient (C_c), Texture ratio (T) and discharge were considered (Kothyari and Garde 1991; Kothyari 2006; Garde 2006) for the analysis. Discharge estimated with the available data of SRSAC, Jodhpur is considered as a dependent variable and the other 10 are considered as independent variables. These were normalized for further analysis as explained earlier.

Feed Forward with Back Propagation

Feed Forward with Back Propagation algorithm is employed with network architecture consisting of 10 inputs, one hidden layer with 6 neurons and one output. The Levenberg–Marquardt (LM) optimization technique (Kisi 2008) is used for training the network. Activation function employed is tansigmoid. Parameters used in training are: learning rate 0.55, maximum number of epochs 1000 and goal (error tolerance) to be achieved is 0.0001. Out of 25 datasets, 17 were used for training whereas 8 were used for testing. These were finalized after extensive trial and error and suggestions from various experts. Network architecture for FFBP is presented in Figure 2.

Figure 15 presents epochs versus squared error values. It is observed that goal is achieved at epoch 23. Figure 16(a) and 16(b) present scatter plots between observed discharge and estimated discharge for the chosen network for 17 training data sets and 8 testing data sets. It is observed that R^2 value for training and testing are 0.998 and 0.485 with the chosen set of parameters. Table 11(a) presents weights of the connection links between the input layer to hidden layer whereas Table 11(b) presents similar values between hidden and output layers for the chosen network. It is observed from Tables 11(a) and 11(b) that there is no trend of the weights as expected as these are nothing but coefficients that can be multiplied with parameters for desired outcome. For example 1.8364 is the coefficient for parameter D_d. Mathematical expression which gives weighted input from 10 inputs (from input layer) to node 1 in hidden layer is as follows: $1.8364\ D_d + 0.8452\ R_b + 1.9207\ F_u - 0.3912\ L_o - 0.752\ R_f - 1.8567\ B_s - 0.2862\ R_e + 0.2767\ R_c - 0.7317\ C_c - 0.6125\ T$. Similarly other mathematical expressions can be formulated based on values of the weights. Table 12 presents comparison of observed discharge and those estimated from feed forward network. Sum of squared deviation of all microwatersheds was 1.0203. Various forms of extensive sensitivity analysis is also performed. It is observed (results not presented) that (i) effect of learning rates on squared error value is significant even though no fixed trend is observed (ii) number of epochs required to reach the target goal are much lesser than the permitted epochs (iii) effect of number of hidden layers and hidden neurons are significant on squared error values (iv) effect of activation functions such as tansigmoid and log sigmoid is also significant on squared error values.

Radial Basis function

Radial Basis Function methodology is used with error goal of 0.0001. Out of 25 datasets, 17 were used for training whereas 8 were used for testing. Spread constant chosen for the study is 3. Figure 17 presents epochs versus squared error values. It is observed that maximum goal is achieved at epoch 15. Figure 18(a) and 18(b) present scatter plots between observed discharge and estimated discharge for 17 training data sets and 8 testing data sets with the chosen parameters. It is observed that R^2 value for training and testing data sets are 1.000 and 0.58 with the chosen parameters.

Table 12 presents comparison of observed discharge and those estimated from Radial Basis Function network. In this case, sum of squared deviation of observed and estimated discharge values for all microwatersheds were 0.6075. It is observed that error is almost half that by FFBP. Effect of spread constant on squared error value is significant but no fixed trend is observed (either decreasing or increasing).

Multiple Linear Regression

Multiple Linear Regression is used for comparing with solutions of FFBP and RBF. Good R^2 value of 0.6277 is observed in this case. Discharge values predicted by Multiple Linear Regression are also presented in Table 12 and in this case, sum of squared deviation value is found to be 0.4976. It is less than those of FFBP and RBF. Mathematical expression for Multiple Linear Regression is as follows:

$$Q = -0.2705D_d - 0.0409\, R_b + 0.0151\, F_u - 0.1616\, L_o - 0.1463\, R_f + 0.4043\, B_s - 0.5234\, R_e - 0.0325\, R_c + 0.0021\, C_c + 0.2164\, T + 0.7149 \quad (9)$$

It is observed that discharge is having reasonably good correlation with 10 morphological parameters for the chosen set of parameters. However, more analysis is necessary for further inferences which are targeted for future research.

Regarding suitability among three prediction techniques namely, Regression, Feed forward network using back propagation (FFBP), Radial basis function (RBF) network: Even though square error value for multiple regression scenario is less (0.4976) as compared to other two i.e., 1.0203, 0.6076 respectively for FFBP and RBF (less square error value for RBF as compared to FFBP), estimated values in the regression scenario are negative for some data sets (8 and 10) for the present planning problem (Table 12). Keeping this aspect in view, regression is not considered for suitability scenario. Among other two i.e., FFBP and RBF; RBF is found to be suitable due to its reasonable R^2 values both for training and testing (1 and 0.58) as compared to FFBP (0.998 and 0.485).

Keeping this aspect in view Radial basis function approach is found to be suitable for prediction purpose.

Inferences: Prediction Aspects

Three prediction techniques, namely, Feed Forward with Back Propagation, Radial Basis Function and Multiple Linear Regression were employed for establishing relationship between discharge and Drainage density (D_d), Bifurcation ratio (R_b), Stream frequency (F_u), Length of overland flow (L_o), Form factor (R_f), Shape factor (B_s), Elongation ratio (R_e), Circulatory ratio (R_c), Compactness coefficient (C_c), Texture ratio (T) for Kherthal watershed. The results of these methodologies are analyzed and compared. The following inferences are drawn from the study:

1. Radial basis function approach is found to be suitable for prediction purpose. However, more analysis is necessary for further inferences. This includes analysis of more microwatersheds.
2. It is observed from the analysis that correlation coefficients for training both in FFBP and RBF are 0.998 and 1 whereas these are 0.485 and 0.58 in case of testing. The lesser values in testing may be due to heterogeneity nature of microwatersheds and less data available for testing.
3. Prediction methodology based on morphometric analysis can be extended for wider applications of watershed development and management.

9. CONCLUSIONS/ RECOMMENDATIONS

Various algorithms are applied to case study of Kherthal watershed, Rajasthan for classification, namely, Kohonen Neural Networks, Fuzzy Cluster Analysis and Cluster

Analysis; and Feed Forward with Back Propagation, Radial Basis Function networks and Multiple Linear Regression for prediction. Applicability of various cluster validation indices is also studied.

It is inferred from results of the above studies, along with inferences of each perspective, that the outcome varied from algorithm to algorithm with the chosen set of parameters, data availability, but the methodology remains the same which is the main objective and focus of the present study. Cluster Analysis methodology can be used as the basis for clustering microwatersheds whereas Radial basis function approach is found to be suitable for prediction purpose.

The proposed neural network and other related methodologies enhances the decision making capability of the water resources planners as classification and prediction can be done with more accuracy with less time. The methodology has academic as well as practical value and would benefit researchers working in water resources and allied fields.

10. HOW DO THE CONCLUSIONS/ RECOMMENDATIONS COMPARE WITH CURRENT THINKING

Special training on Artificial Neural Networks and other related methodologies is necessary for the field experts to implement the methodology in a meaningful way. Moreover some persuasion is also required to the field experts about the validity and potentiality of the methodology to tune their thinking with the outcome of the study.

In addition, workshops can be conducted so that effective knowledge dissemination can be made to enable field experts/ scientists/ academicians to understand the idea behind the methodology and its potential advantages. This also paves the way to put the results into practical use.

11. FIELD TESTS CONDUCTED

Field tests were not conducted. However, number of field visits are made to (i) interact with various stake holders in the Kherthal watershed (ii) understand their adopted conservation measures (iii) understand constraints faced by them. This helped to appropriately interpret the various characteristics of the watershed. It is inferred from these discussions and observation from the field visits that all the 25 microwatersheds may not require similar conservation measures/ treatment which may necessitate to formulate homogeneous groups of microwatersheds to handle the problem accordingly.

12.SOFTWARE GENERATED, IF ANY: No

13.POSSIBILITIES OF ANY PATENTS/COPYRIGHTS. IF SO, THEN ACTION TAKEN IN THIS REGARD: No

14. SUGGESTIONS FOR FURTHER WORK

Further studies can be extended with more analysis of RBF as well as exploring Quadratic regression and other improvements that may be useful for Artificial Neural Networks (Zhang, 2007). Fuzzy Cluster Analysis can be explored as the basis for cluster validation techniques in further studies, along with cost analysis and homogeneity measures.

If more data is available, two phase approach may be explored where classification algorithm outcome can be used to determine homogeneous groups and prediction algorithms can be used to formulate groupwise relationships between parameters of interest to ensure regional homogeneity and more accurate prediction of run-off.

Publications emanated from the present studies are presented below whereas Annexure-2 presents list of references whereas Annexutre-3 presents acknowledgements.

PUBLICATIONS

Raju KS, Nagesh Kumar, D (2007) Classification of indian meteorological stations using cluster and fuzzy cluster analysis, and Kohonen artificial neural networks. Nordic Hydrology 38: 303–314

Raju KS, Nagesh Kumar, D (2009) Classification of microwatersheds based on morphological data, Journal of Hydro Environment Research (under review)

Raju KS, Nagesh Kumar, D (2006) Taguchi Methodology for Multicriterion Decision Making in Irrigation Planning. An International Perspective on Environmental and Water Resources 2006, EWRI of ASCE, December 18-20, 2006, New Delhi (Full paper in CD-ROM).

Raju KS, Kalra NK, Bhatia N (2008) Classification of microwatersheds using Kohonen Neural Networks and Cluster Analysis: A Case Study in Rajasthan. International Convention on Water Resources Development and Management (ICWRDM 2008), October 23-24, 2008, BITS, Pilani, pp.67.

Table 1. Salient parameters of the 25 microwatersheds of Kherthal Watershed

Parameter / No.	Area (km^2)	Perimeter (km)	Channel Slope (%)	Mean Basin Elevation (m)	Basin Length (km)
1	2.750	7.767	1.000	345.000	2.331
2	50.500	38.772	1.500	458.500	12.173
3	12.500	17.067	3.540	546.500	5.508
4	6.750	13.163	7.120	587.500	3.881
5	3.000	8.017	4.840	539.500	2.449
6	4.250	9.277	25.750	732.500	2.984
7	4.000	11.993	19.310	687.500	2.883
8	0.500	3.077	8.000	490.000	0.885
9	2.000	6.511	16.600	641.000	1.945
10	0.750	4.344	19.630	632.000	1.114
11	1.750	10.895	38.000	615.000	1.803
12	10.500	21.938	15.720	697.000	4.988
13	3.200	10.759	12.500	550.000	2.540
14	1.300	6.154	10.660	520.000	1.523
15	2.500	6.949	14.400	620.000	2.208
16	1.500	6.630	33.000	527.500	1.652
17	4.250	11.365	7.000	635.000	2.984
18	14.000	19.107	9.920	772.000	5.874
19	2.500	7.318	4.000	480.000	2.208
20	6.750	15.897	1.160	435.000	3.881
21	2.250	7.144	4.000	420.000	2.080
22	1.130	5.388	14.750	449.000	1.406
23	4.500	16.197	2.800	375.000	3.083
24	13.550	19.553	12.620	462.500	5.766
25	2.250	7.910	3.500	435.000	2.080

Table 2. Number of streams of each microwatershed and length

Parameter No	No. of streams in first order	second order	third order	fourth order	Total number of streams	Total length (km)
1	1	--	--	--	1	1.05
2	161	81	37	40	319	171.31
3	55	24	24	3	106	47.93
4	57	23	16	9	105	34.60
5	25	9	10	--	44	13.73
6	24	13	3	1	41	11.46
7	49	28	6	6	89	31.94
8	2	--	--	--	2	1.113
9	15	8	2	--	25	8.95
10	7	2	1	--	10	3.32
11	11	5	--	--	16	6.94
12	83	34	22	--	139	49.34
13	21	9	4	--	34	12.71
14	8	3	--	--	11	4.50
15	16	7	5	--	28	11.16
16	7	3	--	--	10	4.19
17	20	7	12	--	39	18.28
18	67	37	9	12	125	56.99
19	18	10	4	--	32	12.17
20	23	10	8	--	41	23.75
21	6	6	--	--	12	7.61
22	2		--	--	2	1.815
23	9	5	1	--	15	8.906
24	36	20	6	--	62	38.13
25	13	6	--	--	19	8.61

Table 3. Description of morphometric and other related parameters

Parameters	Unit	Formulae	Analysis for microwatershed 3
Basin length	Km	$L_b = 1.312A^{0.568}$	5.508
Drainage Density	Km^{-1}	$D_d = \frac{L}{A}$	3.834
Stream Frequency	Km^{-2}	$F_u = \frac{N}{A}$	8.48
Length of overland flow	Km	$L_o = \frac{0.5}{D_d}$	0.130
Bifurcation ratio	No unit	$R_b = \frac{N_u}{N_{u+1}}$	2.292,1,8 Average is 11.292/3=3.764
Form factor	No unit	$R_f = \frac{A}{L_b^2}$	0.4120
Shape factor	No unit	$B_s = \frac{L_b^2}{A}$	2.4272
Elongation ratio	No unit	$R_e = 1.128\frac{A^{0.5}}{L_b}$	0.7241
Circulatory ratio	No unit	$R_c = 12.57\frac{A}{P^2}$	0.5394
Compactness Coefficient	No unit	$C_c = 0.2821\frac{P}{A^{0.5}}$	1.3617
Texture ratio	Km^{-1}	$T = \frac{N_1}{P}$	3.222

Where, for microwatershed number 3,

A	=	Area of the microwatershed (km^2)	12.5
P	=	Perimeter (km)	17.067
L	=	Total length of all streams of all orders (Km)	47.9332
N	=	Total number of streams	106
N_1	=	Total number of first order streams	55
N_u	=	No. of streams of order, u	55, 24, 24, 3
N_{u+1}	=	No. of streams of next higher order, u	24, 24, 3

Table 4. Morphological parameters of microwatersheds in Kherthal watershed

Parameter / No.	D_d (Km/Km^2)	R_b (No unit)	F_u (No. of. streams/ km^2)	L_o (Km)	R_f (No unit)	B_s (No unit)	R_e (No unit)	R_c (No unit)	C_c (No unit)	T (Km^{-1})
1	0.384	0.000	0.364	**1.303**	0.506	1.975	0.803	0.573	1.321	0.129
2	3.392	1.701	6.337	0.147	**0.341**	2.934	**0.658**	0.422	1.539	4.152
3	3.834	**3.764**	8.480	0.130	0.412	2.427	0.724	0.539	1.362	3.223
4	5.126	1.898	15.556	0.098	0.448	2.232	0.755	0.490	1.429	**4.330**
5	4.577	1.839	14.667	0.109	0.500	1.999	0.798	0.587	1.306	3.118
6	2.696	3.060	9.647	0.185	0.477	2.096	0.779	0.621	1.269	2.587
7	**7.985**	2.472	**22.250**	0.063	0.481	2.078	0.782	0.350	1.692	4.086
8	2.226	0.000	4.000	0.225	0.638	**1.566**	0.901	0.664	**1.228**	0.650
9	4.477	2.938	12.500	0.112	0.529	1.892	0.820	0.593	1.299	2.304
10	4.436	2.750	13.333	0.113	0.604	1.655	0.877	0.500	1.415	1.612
11	3.969	2.200	9.143	0.126	0.538	1.857	0.828	**0.185**	2.323	1.010
12	4.700	1.993	13.238	0.106	0.422	2.370	0.733	0.274	1.910	3.783
13	3.974	2.292	10.625	0.126	0.496	2.016	0.794	0.347	1.697	1.952
14	3.464	2.667	8.462	0.144	0.561	1.784	0.845	0.432	1.522	1.300
15	4.464	1.843	11.200	0.112	0.513	1.950	0.808	0.651	1.240	2.303
16	2.798	2.333	6.667	0.179	0.550	1.819	0.836	0.429	1.527	1.056
17	4.303	1.720	9.176	0.116	0.477	2.096	0.779	0.414	1.555	1.760
18	4.071	2.224	8.929	0.123	0.406	2.465	0.719	0.482	1.441	3.507
19	4.870	2.150	12.800	0.103	0.513	1.950	0.808	0.587	1.306	2.460
20	3.520	1.775	6.074	0.142	0.448	2.232	0.755	0.336	1.726	1.447
21	3.385	1.000	5.333	0.148	0.520	1.922	0.814	0.554	1.344	0.840
22	1.607	0.000	1.770	0.311	0.571	1.750	0.853	0.489	1.430	0.371
23	1.979	3.400	3.333	0.253	0.473	2.112	0.776	0.216	2.154	0.556
24	2.814	2.567	4.576	0.178	0.408	2.454	0.720	0.446	1.498	1.841
25	3.829	2.167	8.444	0.131	0.520	1.922	0.814	0.452	1.488	1.643
Nature	Max	Max	Max	Max	Min	Min	Min	Min	Min	Max
Value	7.985	3.764	22.250	1.303	0.341	1.566	0.658	0.185	1.228	4.330

Table 5. Normalized morphological parameters of microwatersheds in Kherthal watershed

Parameter / No.	D_d	R_b	F_u	L_o	R_f	B_s	R_e	R_c	C_c	T
1	0.048	0.000	0.016	1.000	1.486	1.261	1.219	3.092	1.076	0.030
2	0.425	0.452	0.285	0.113	1.000	1.873	1.000	2.279	1.254	0.959
3	0.480	1.000	0.381	0.100	1.209	1.549	1.100	2.911	1.109	0.744
4	0.642	0.504	0.699	0.075	1.315	1.425	1.147	2.643	1.164	1.000
5	0.573	0.489	0.659	0.084	1.468	1.276	1.212	3.166	1.064	0.720
6	0.338	0.813	0.434	0.142	1.400	1.338	1.183	3.349	1.034	0.597
7	1.000	0.657	1.000	0.048	1.412	1.327	1.188	1.886	1.378	0.943
8	0.279	0.000	0.180	0.172	1.873	1.000	1.369	3.582	1.000	0.150
9	0.561	0.780	0.562	0.086	1.551	1.207	1.246	3.200	1.058	0.532
10	0.556	0.731	0.599	0.086	1.773	1.057	1.331	2.696	1.153	0.372
11	0.497	0.585	0.411	0.097	1.580	1.186	1.257	1.000	1.893	0.233
12	0.589	0.530	0.595	0.082	1.238	1.513	1.113	1.480	1.556	0.874
13	0.498	0.609	0.478	0.097	1.455	1.287	1.206	1.875	1.382	0.451
14	0.434	0.708	0.380	0.111	1.645	1.139	1.283	2.329	1.240	0.300
15	0.559	0.490	0.503	0.086	1.505	1.245	1.227	3.512	1.010	0.532
16	0.350	0.620	0.300	0.137	1.613	1.161	1.270	2.315	1.244	0.244
17	0.539	0.457	0.412	0.089	1.400	1.338	1.183	2.232	1.267	0.406
18	0.510	0.591	0.401	0.094	1.191	1.573	1.091	2.601	1.174	0.810
19	0.610	0.571	0.575	0.079	1.505	1.245	1.227	3.166	1.064	0.568
20	0.441	0.472	0.273	0.109	1.315	1.425	1.147	1.812	1.406	0.334
21	0.424	0.266	0.240	0.113	1:527	1.227	1.236	2.990	1.094	0.194
22	0.201	0.000	0.080	0.239	1.677	1.117	1.295	2.640	1.165	0.086
23	0.248	0.903	0.150	0.194	1.389	1.348	1.179	1.163	1.755	0.128
24	0.352	0.682	0.206	0.136	1.196	1.566	1.094	2.404	1.221	0.425
25	0.480	0.576	0.380	0.100	1.527	1.227	1.236	2.439	1.212	0.380

Table 6. Cluster validation indices computed based on Cluster Analysis as classification algorithm

Index / No. of groups	Davies Bouldin	Dunn's	Silhouette
3	1.0229	1.333	0.24669
4	0.92585	1.6704*	0.2805
5	0.95626	1.34	0.30145
6	0.84115*	1.3775	0.33577*

*Optimal value of respective indices

Table 7. Classification of microwatersheds by FCA, KNN and CA

Technique / Group	FCA	KNN	CA	Common micro watersheds among			
				All methods	FCA and KNN	FCA and CA	KNN and CA
1	11,23 **(2)**	14,16,17,25 **(4)**	11,12,23**(3)**	Nil	Nil	2	Nil
2	5,6,8,9,15,19,21(7)	5,6,9,15,19 **(5)**	5,6,9,10,15,19**(6)**	5	5	5	5
3	7,12,13,17,20 **(5)**	11,12,13,20,23 **5)**	7,13,14,16,17,25**(6)**	1	3	3	1
4	2,3,4,18,24 **(5)**	2,3,4,7,18,24**(6)**	2,3,4,18,20,24**(6)**	5	5	5	5
5	1,10,14,16,22,25**(6)**	1,8,10,21,22**(5)**	1,8,21,22**(4)**	2	3	2	4

Table 8. Final weights for Kohonen neural networks

Parameter / Group	D_d	R_b	F_u	L_o	R_f	B_s	R_e	R_c	C_c	T
1	0.4363	0.6098	0.3460	0.1134	1.4875	1.2760	1.2177	2.3420	1.2377	0.3518
2	0.5274	0.6337	0.5446	0.0954	1.4824	1.2657	1.2173	3.2737	1.0468	0.5918
3	0.4625	0.6204	0.3896	0.1148	1.3945	1.3527	1.1799	1.4661	1.5982	0.4122
4	0.6012	0.6371	0.5408	0.0875	1.2217	1.5536	1.1035	2.4671	1.2144	0.8828
5	0.2965	0.1886	0.2152	0.3266	1.6642	1.1343	1.2888	3.0071	1.0964	0.1626

Table 9 . Occupancy of microwatersheds in each group and corresponding squared error values for various learning rates and epochs for KNN

Epochs / Learning rate	1000	3000	5000
0.01	4,5,5,6,5 (6.4200)	5,5,5,5,5 (6.2541)	5,5,5,5,5 (6.2544)
0.1	5,4,5,5,6 (6.4085)	5,4,5,5,6 (6.4069)	5,4,5,5,6 (6.3229)
0.2	5,5,4,5,6 (6.3390.	5,4,5,6,5 (6.4257)	6,5,4,5,5 (6.8861)
0.3	4,5,5,5,6 (6.9451)	5,5,5,6,4 (6.7156)	4,6,4,4,7 (6.0519)
0.4	6,4,6,4,5 (7.3714)	4,5,6,5,5 (6.0366)	5,4,5,6,5 (6.6510)
0.5	6,5,6,4,4 (7.0117)	5,5,5,5,5 (7.3007)	6,4,4,6,4 (8.8679)
0.6	6,4,4,4,7 (8.5786)	6,4,6,4,5 (7.0030)	3,6,3,8,5 (7.3460)
0.7	6,4,4,8,3 (7.8704)	4,5,4,7,5(7.9609)	4,6,5,5,5 (7.3937)
0.8	6,4,4,6,5 (8.5335)	8,4,5,3,5 (7.3347)	8,3,4,8,2 (10.0603)
0.9	5,4,5,2,9 (10.906)	7,7,2,7,2 (8.1350)	3,12,6,2,2 (14.376)

Table 10. Membership values of the microwatersheds under each group showing the representative group of each microwatershed

Group / No.	1	2	3	4	**5**	Representative Group
1	0.0857	0.2712	0.1608	0.2028	**0.2795**	5
2	0.0621	0.0924	0.2148	**0.4968**	0.1339	4
3	0.0324	0.2415	0.0988	**0.474**	0.1533	4
4	0.0289	0.1444	0.1163	**0.5817**	0.1288	4
5	0.0115	**0.7758**	0.0342	0.1062	0.0723	2
6	0.0171	**0.7203**	0.0437	0.1239	0.095	2
7	0.1314	0.1076	**0.3563**	0.2275	0.1772	3
8	0.0568	**0.4356**	0.1122	0.1606	0.2348	2
9	0.0095	**0.833**	0.0263	0.0649	0.0663	2
10	0.0358	0.2295	0.1166	0.1399	**0.4783**	5
11	**0.872**	0.0137	0.0616	0.0223	0.0303	1
12	0.2348	0.0577	**0.4396**	0.1472	0.1207	3
13	0.0268	0.0112	**0.8848**	0.0284	0.0488	3
14	0.0239	0.0429	0.1161	0.0631	**0.7541**	5
15	0.0108	**0.8429**	0.0267	0.0623	0.0574	2
16	0.021	0.034	0.0996	0.0511	**0.7941**	5
17	0.0309	0.0448	**0.4601**	0.1385	0.3256	3
18	0.0017	0.0074	0.0081	**0.9736**	0.0091	4
19	0.0042	**0.9182**	0.0124	0.0344	0.0309	2
20	0.0838	0.0264	**0.7217**	0.0681	0.1001	3
21	0.0322	**0.3774**	0.0942	0.1584	0.3378	2
22	0.0678	0.1832	0.1658	0.1571	**0.4261**	5
23	**0.8424**	0.0165	0.0747	0.0283	0.0381	1
24	0.0434	0.0864	0.2232	**0.3914**	0.2557	4
25	0.0067	0.0196	0.0423	0.0366	**0.8948**	5

Table 11 (a). Weights of the connection links between the input layer and the hidden layer

Hidden neuron \ Input neuron	1	2	3	4	5	6	7	8	9	10
1	1.8364	0.8452	1.9207	-0.3912	-0.752	-1.8567	1.2862	0.2767	-0.7317	-0.6125
2	-1.1991	-1.4194	0.7647	0.7325	-2.1676	1.8681	-0.5131	-0.8932	-1.5419	2.1489
3	0.3212	0.4998	-1.2571	-1.6952	-1.1543	-0.3185	-1.1805	1.6843	-1.4059	-0.0534
4	-0.0035	0.0064	-0.363	-1.0122	0.6347	2.1669	0.9306	0.0026	0.9512	1.9294
5	1.5908	-0.2405	1.0933	1.2908	-1.0592	0.612	1.2571	-0.5518	0.43	1.4847
6	0.9228	0.9531	2.8335	-1.4946	-1.1511	-0.0797	-1.6145	-0.4963	-0.2501	0.7713

Table 11(b). Weights of the connection links between the hidden layer and the output layer

Hidden Neuron	Output Neuron
1	-0.3709
2	1.151
3	-0.0726
4	0.7149
5	-0.2004
6	0.394

Table 12. Comparison of normalized discharge values obtained by three prediction methods

Discharge No.	Q_o	Q_{FF}	Q_{RB}	Q_{REGR}	Squared Error Values from Q_o		
					Q_{FF}	Q_{RB}	Q_{REGR}
1	0.077	0.072474	0.077131	0.1035	2.13E-05	1.68E-09	0.000696
2	0.716	0.717441	0.715783	0.7914	2.28E-06	2.16E-08	0.005699
3	0.347	0.308708	0.529139	0.4766	0.001453	0.033237	0.01684
4	0.394	0.397327	0.39405	0.4357	9.10E-06	6.81E-08	0.001716
5	0.171	0.172302	0.172335	0.2586	1.61E-06	1.70E-06	0.00766
6	0.341	0.128015	0.346988	0.3134	0.045213	4.02E-05	0.000744
7	0.303	0.303727	0.303605	0.2788	9.55E-08	3.50E-08	0.000606
8	0.052	0.049309	0.052376	-0.0534*	7.82E-06	7.29E-08	0.002714
9	0.161	0.108357	0.11742	0.1488	0.002771	0.001899	0.000148
10	0.065	0.066562	0.063024	-0.0037*	1.90E-06	4.67E-06	0.004251
11	0.174	0.173992	0.174516	0.1596	1.70E-07	1.25E-08	0.000219
12	0.786	0.542359	0.502082	0.5224	0.059575	0.080858	0.069738
13	0.284	0.282368	0.282399	0.2627	1.33E-06	1.26E-06	0.000432
14	0.115	0.130574	0.121168	0.0970	0.000228	3.23E-05	0.000343
15	0.170	0.02535	0.144825	0.1816	0.02104	0.000654	0.000125
16	0.145	0.119199	0.143293	0.1261	0.00065	1.95E-06	0.000345
17	0.168	0.15348	0.169232	0.2772	0.000214	1.26E-06	0.011903
18	0.524	0.255358	0.411014	0.5278	0.071913	0.012659	1.82E-05
19	0.102	0.098436	0.10156	0.1859	1.32E-05	2.54E-07	0.007031
20	0.124	0.136246	0.125028	0.3629	0.000139	3.25E-07	0.056862
21	0.083	0.040194	0.081491	0.1480	0.00179	1.03E-06	0.004284
22	0.099	0.10648	0.09819	0.0872	6.22E-05	1.65E-07	0.00013
23	0.155	0.152961	0.154163	0.3005	2.63E-06	1.76E-07	0.021303
24	1.000	0.097216	0.308484	0.4753	0.815019	0.478194	0.275308
25	0.090	0.103438	0.087478	0.1828	0.000171	8.22E-06	0.008541
		Sum of squared error values			1.0203	0.607596	0.497659

* Considered as zero for the computation of squared error values

Q_o, is estimated discharge whereas Q_{FF}, Q_{RB}, Q_{REGR} are estimated from Feed Forward, Radial Basis and Regression Analysis respectively.

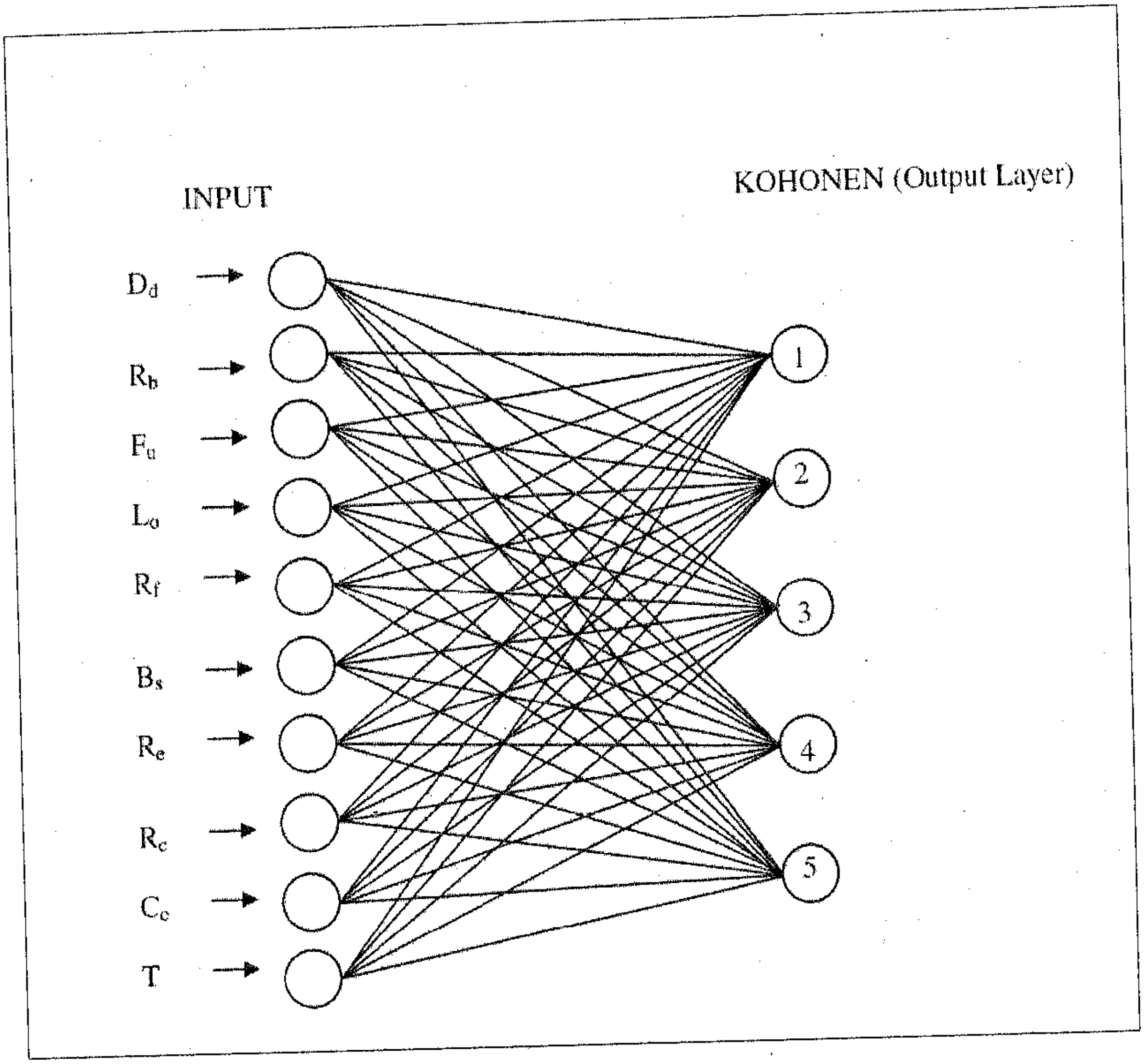

Figure 1. Schematic diagram of Kohonen Neural Networks

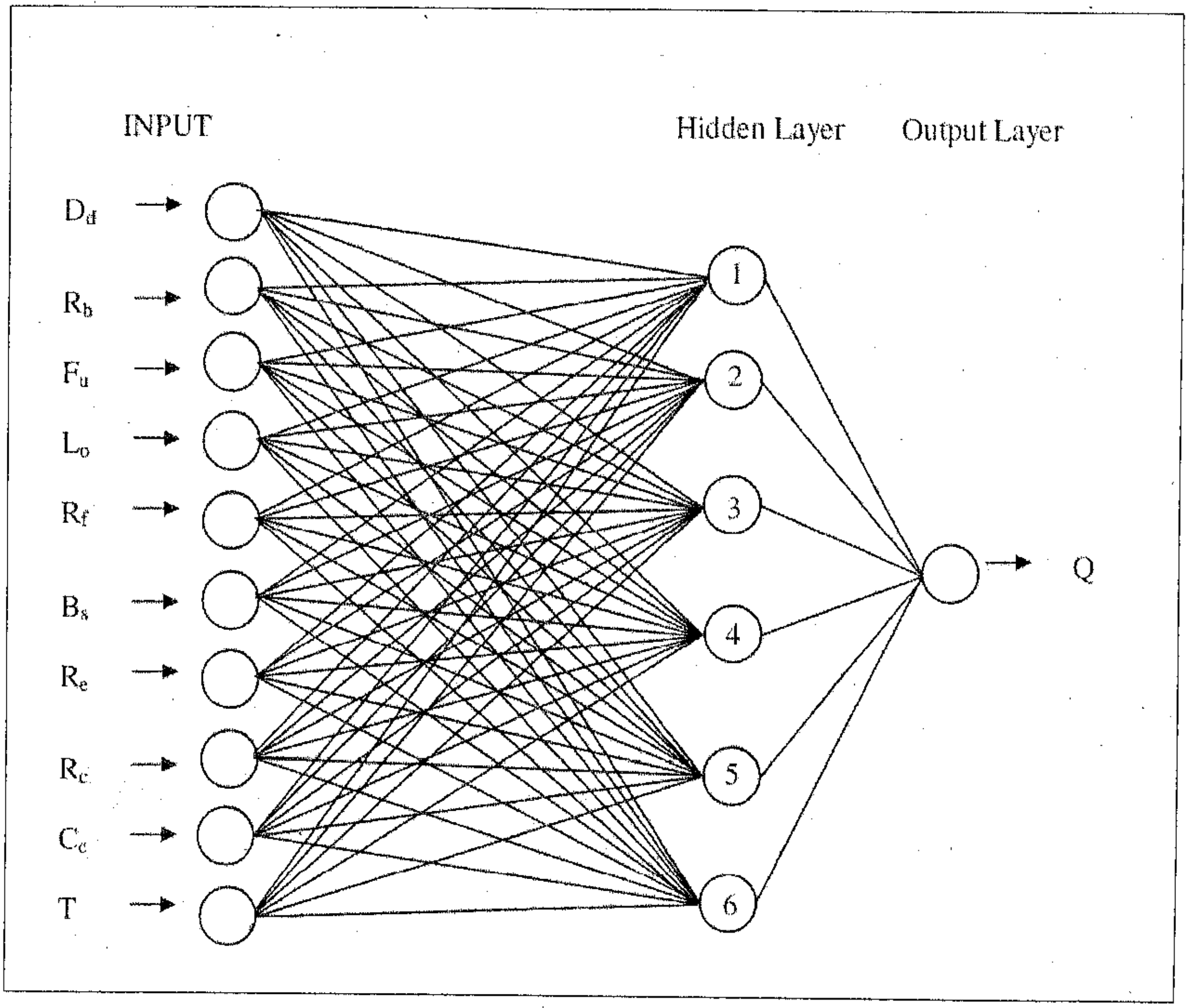

Figure 2. Schematic diagram of Feed Forward with Back Propagation Neural Networks

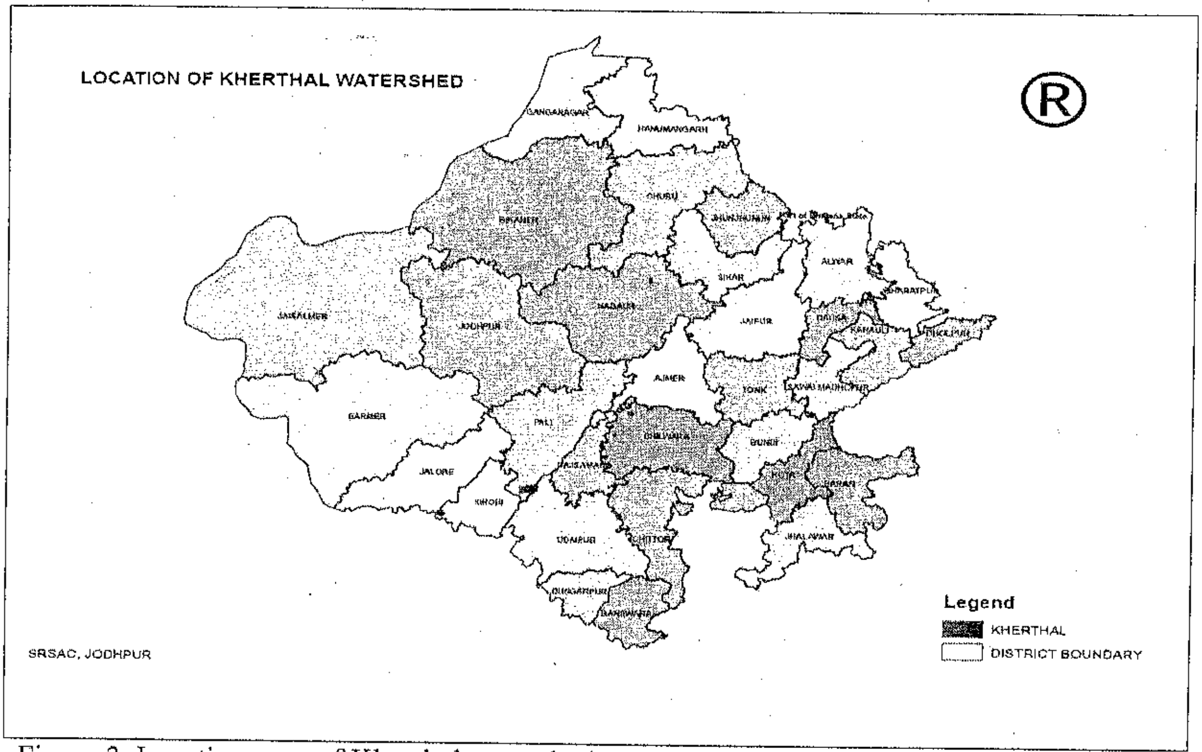

Figure 3. Location map of Kherthal watershed

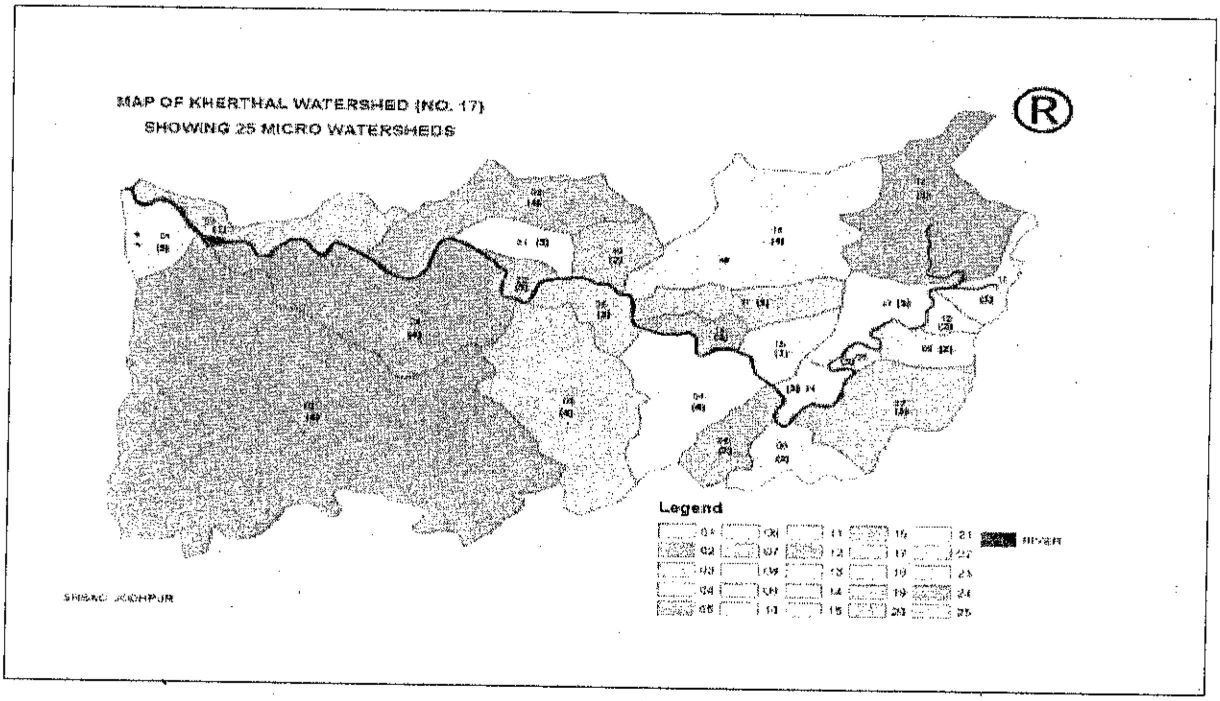

Figure 4. Khethal watershed and its 25 microwatersheds along with corresponding homogeneous regions obtained by cluster analysis methodology

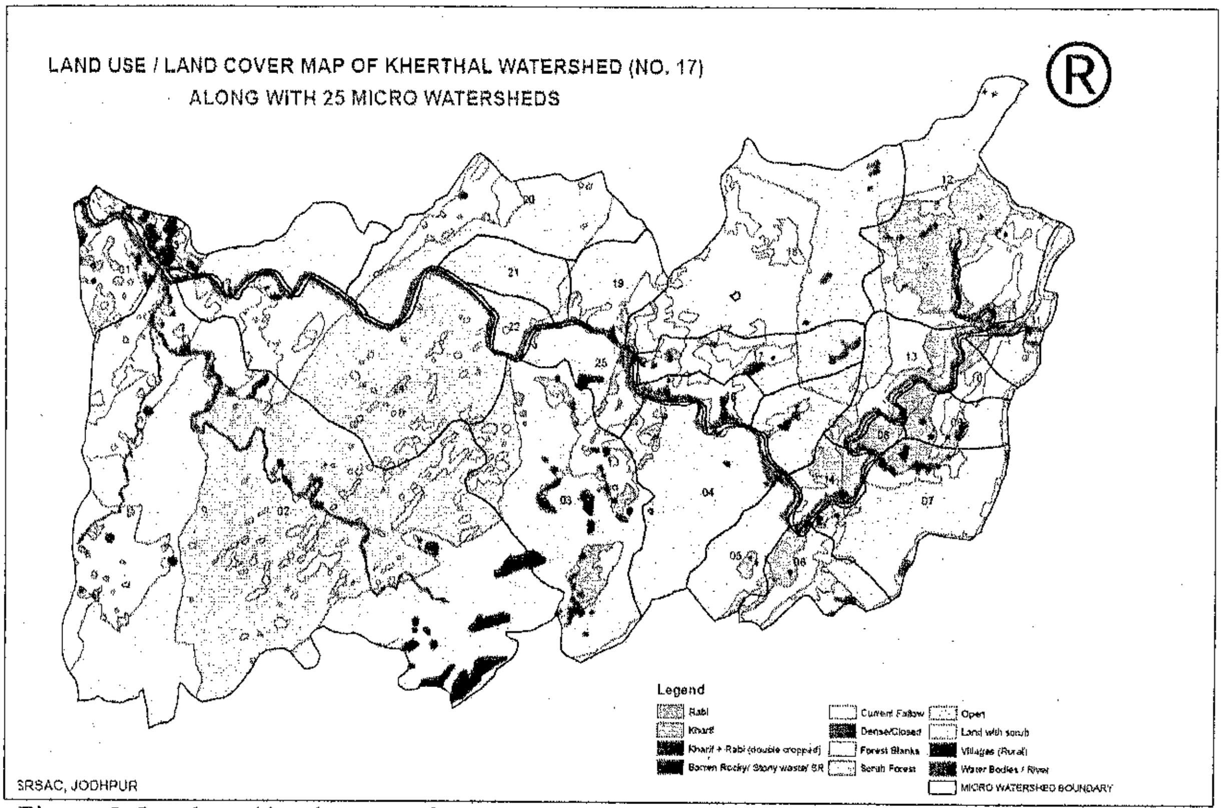

Figure 5. Landuse / land cover of Khethal watershed and its 25 microwatersheds

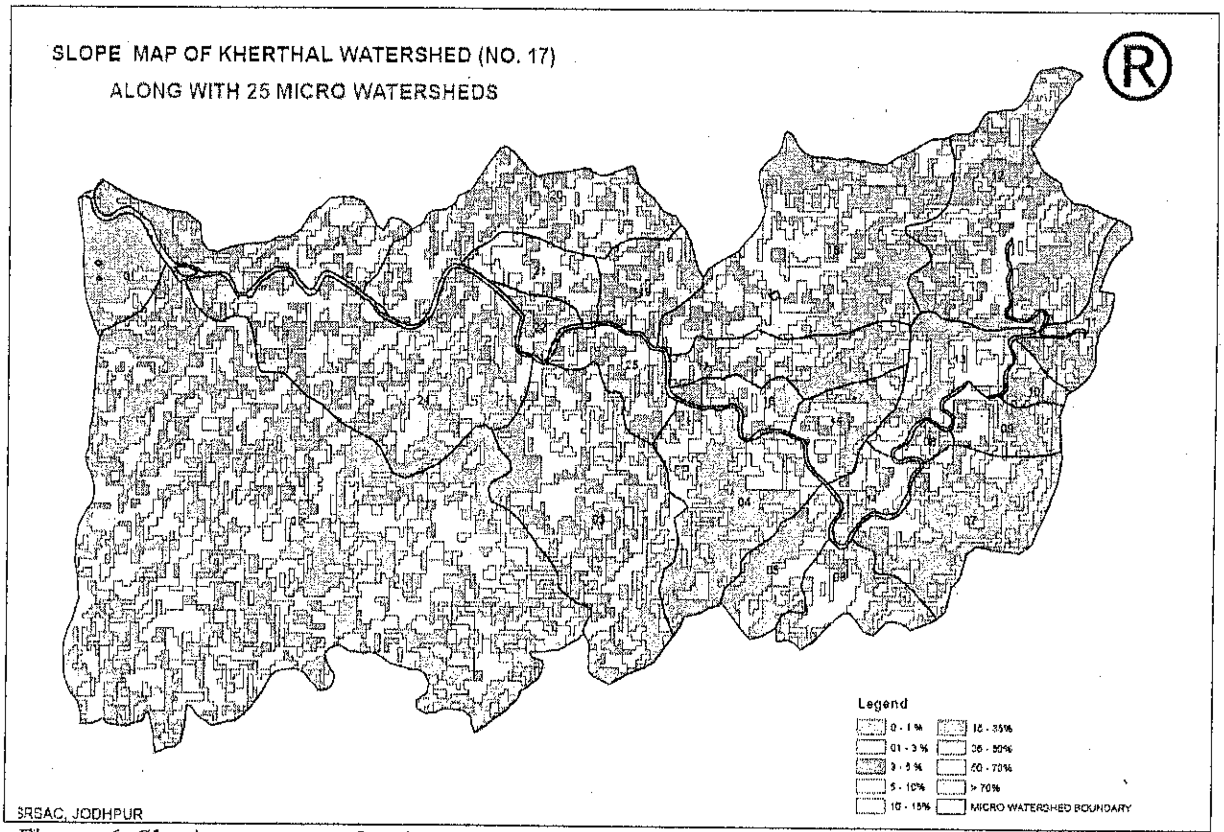

Figure 6. Sloping pattern of microwatersheds

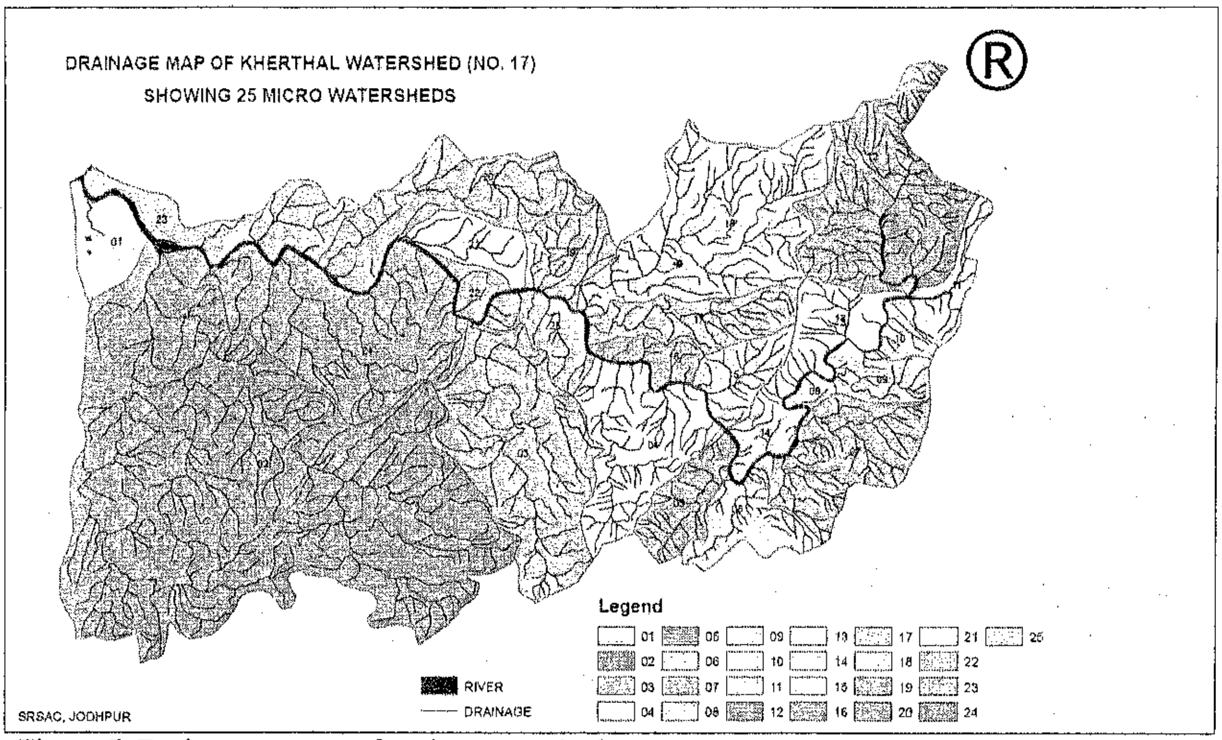

Figure 7. Drainage pattern of microwatersheds

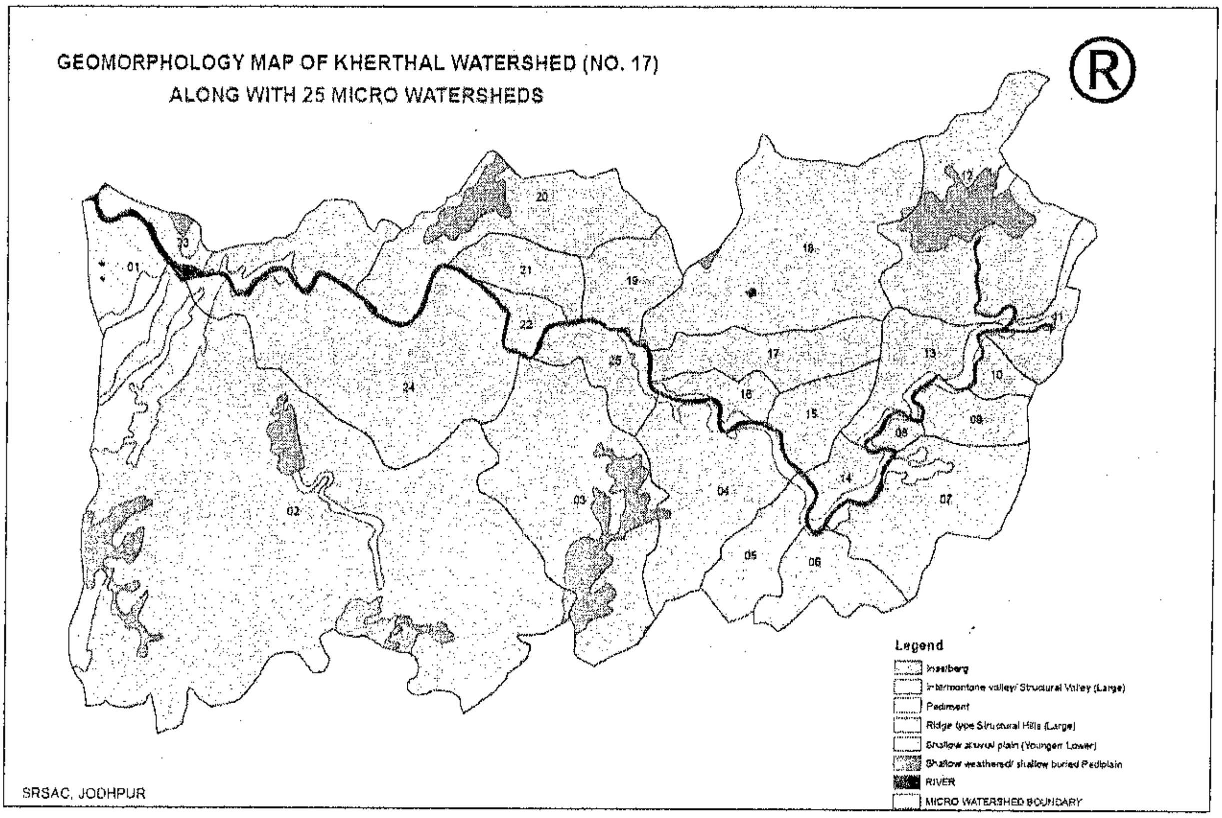

Figure 8. Geomorphology map of microwatersheds

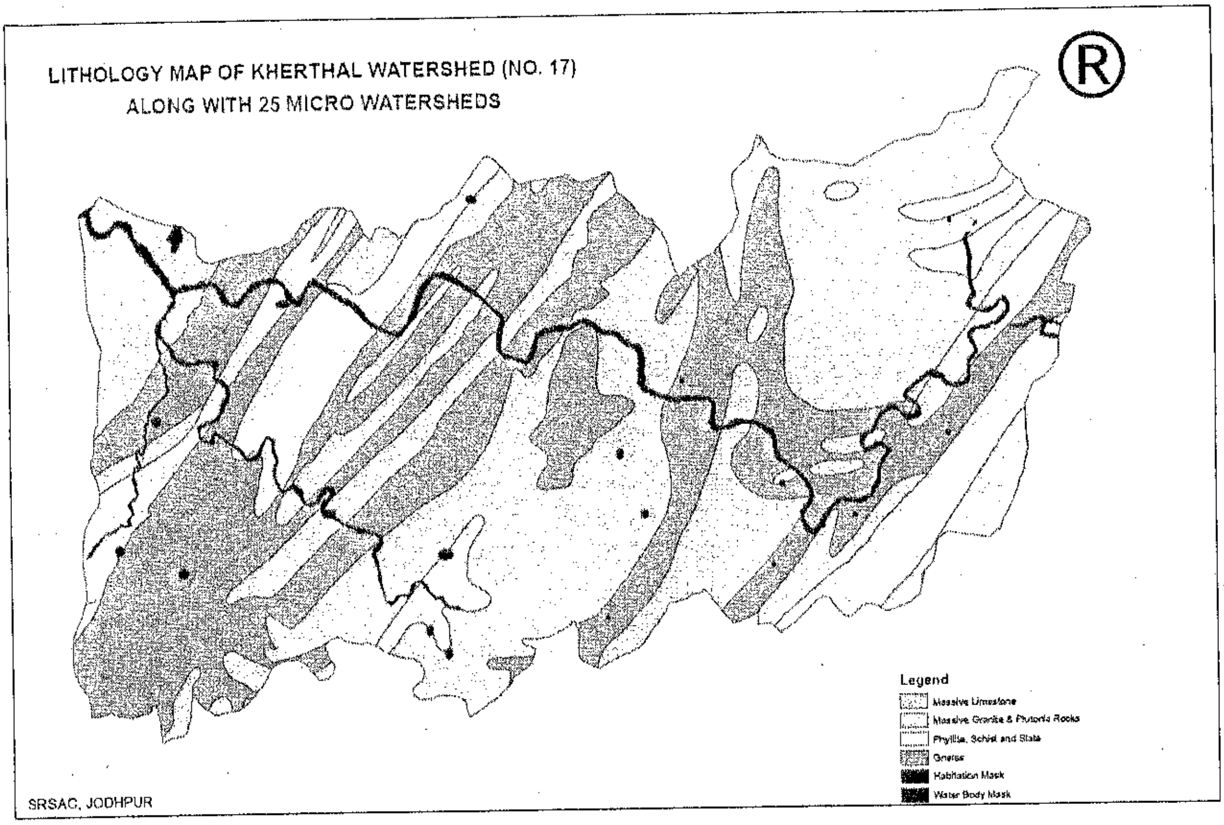

Figure 9. Lithology map of microwatersheds

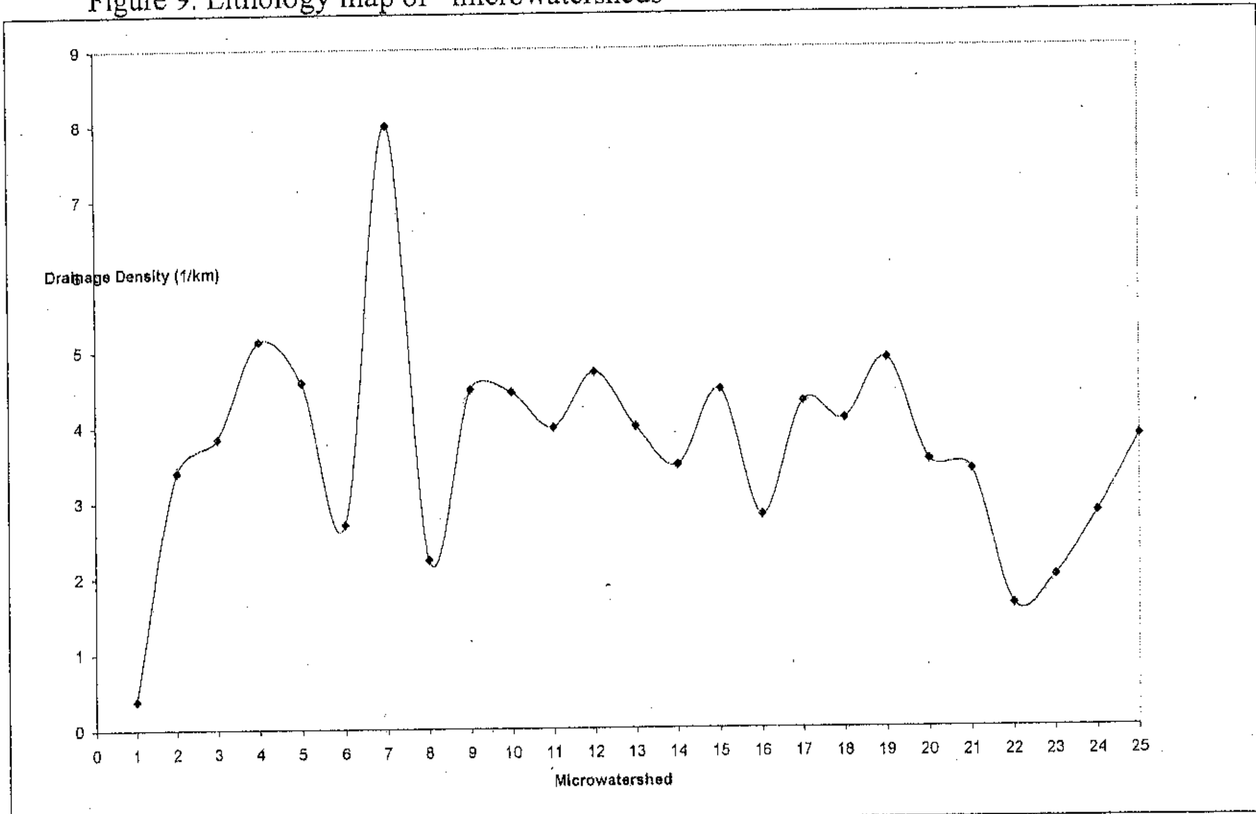

Figure 10. Drainage density values for microwatersheds

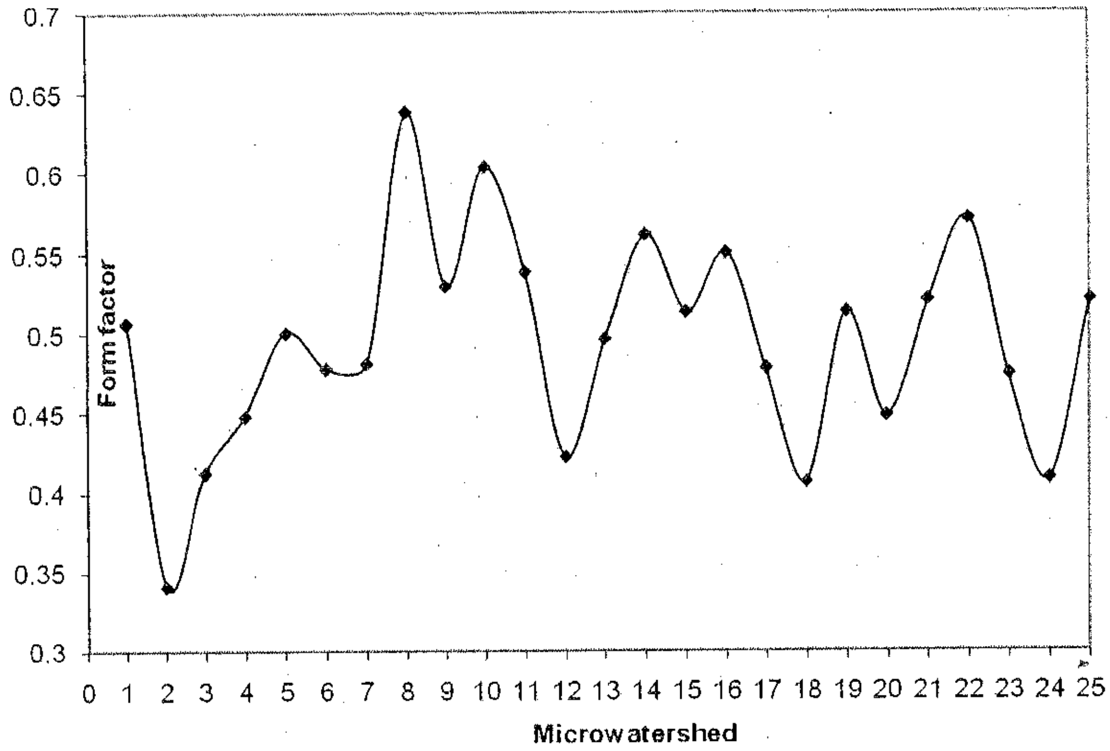

Figure 11. Form factor values for microwatersheds

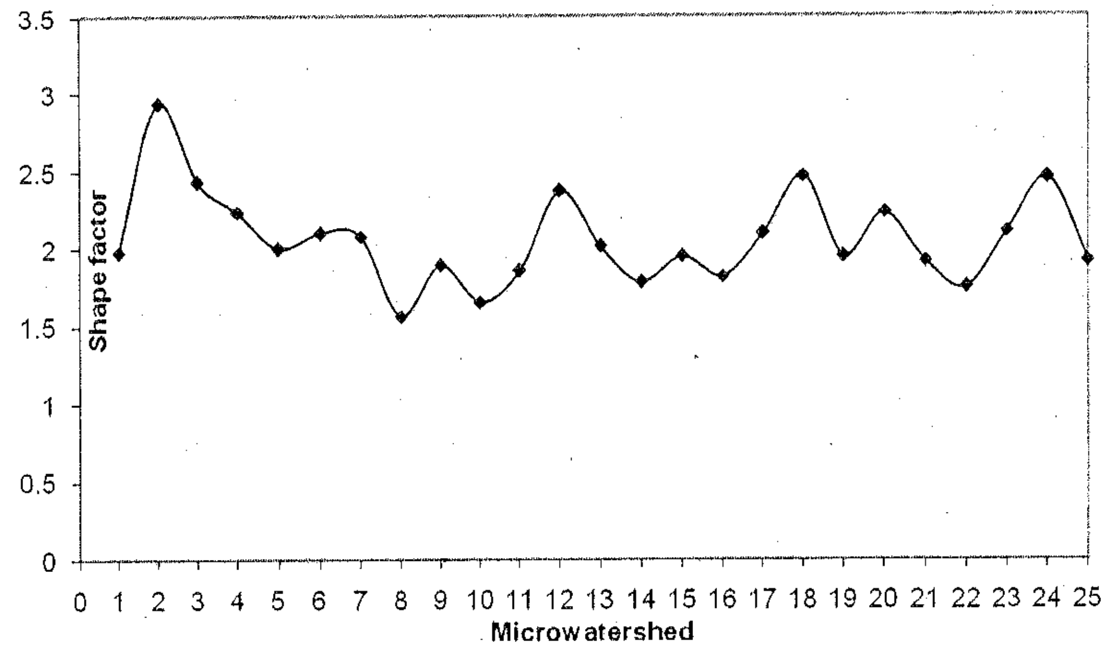

Figure 12. Shape factor values for microwatersheds

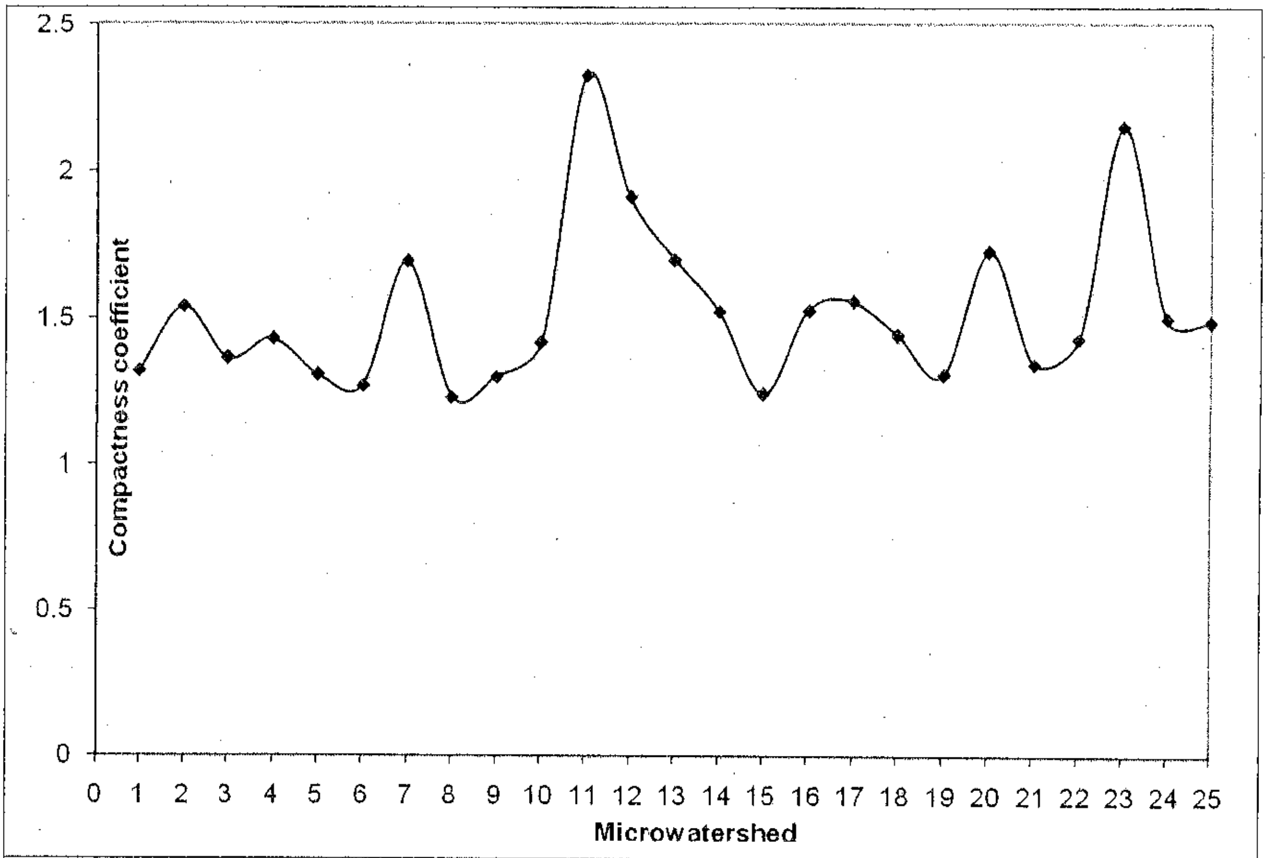

Figure 13. Compactness coefficient values for microwatersheds

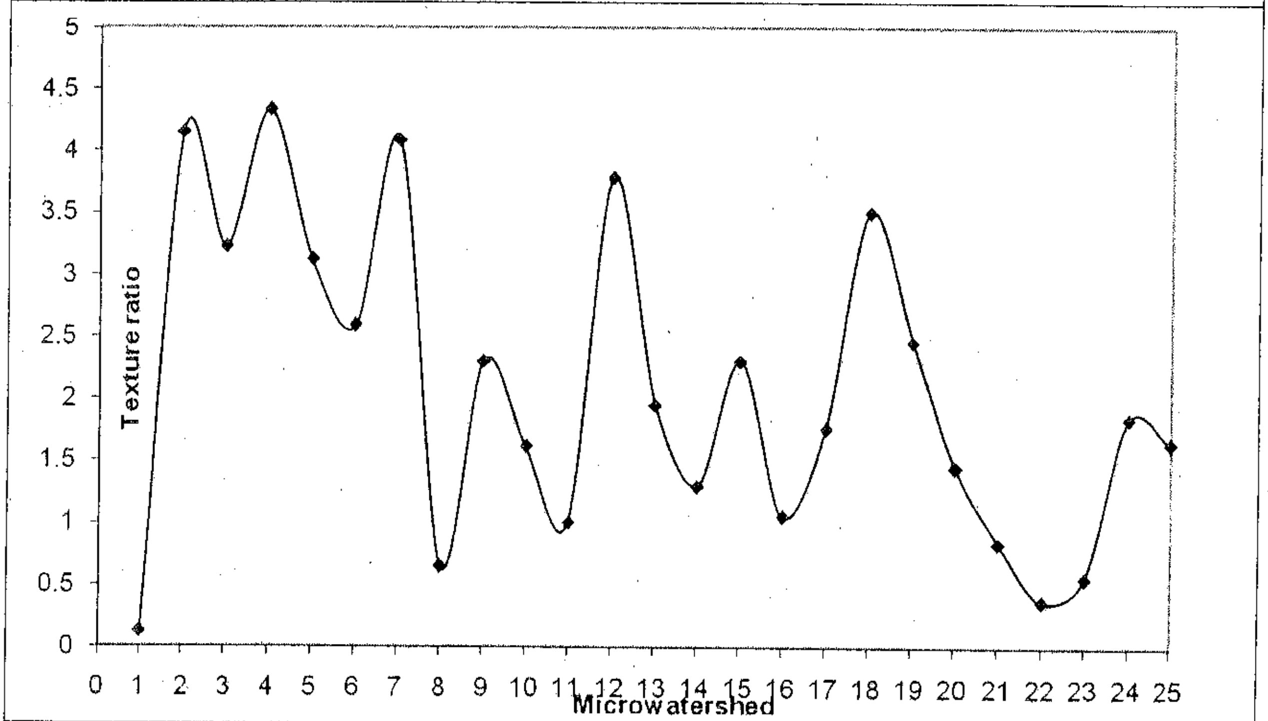

Figure 14. Texture ratio values for microwatersheds

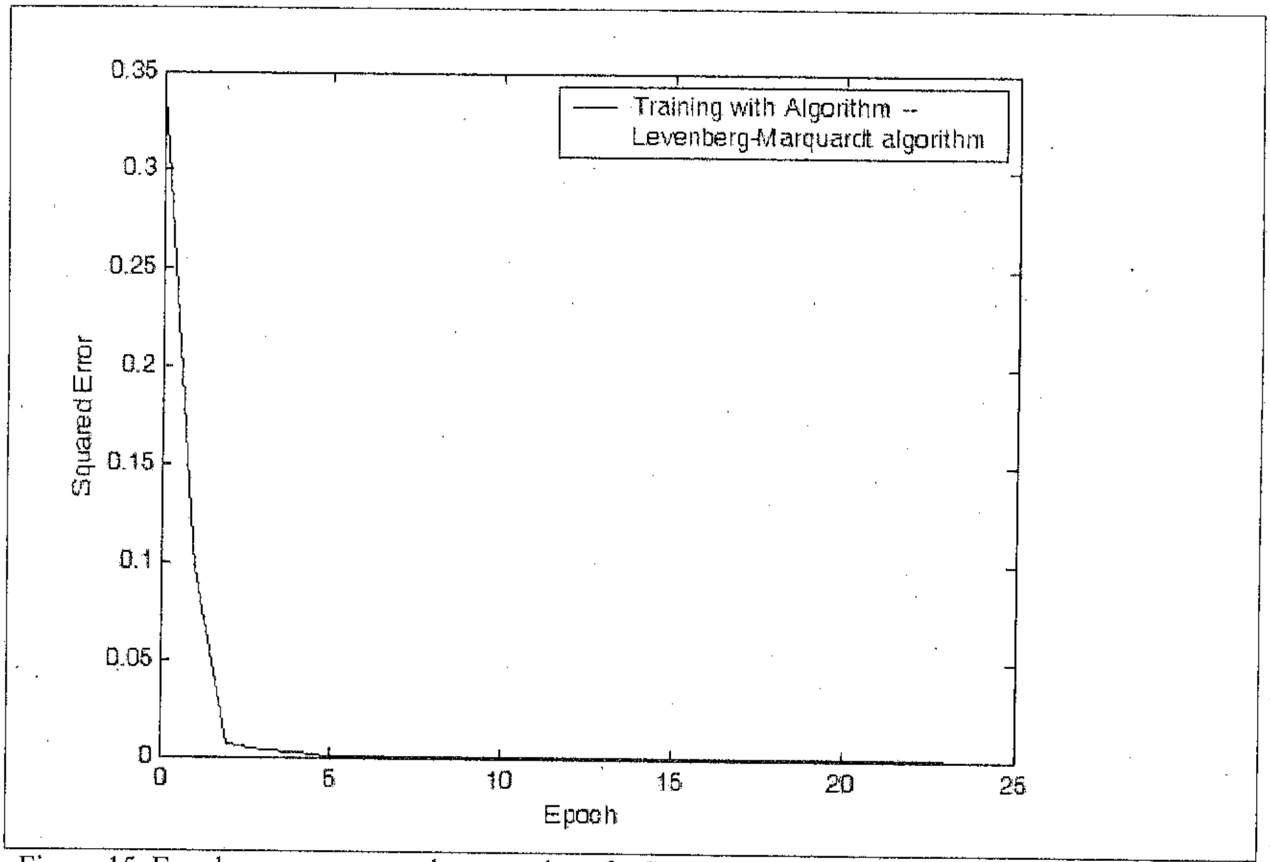

Figure 15. Epochs versus squared error values for FFBP

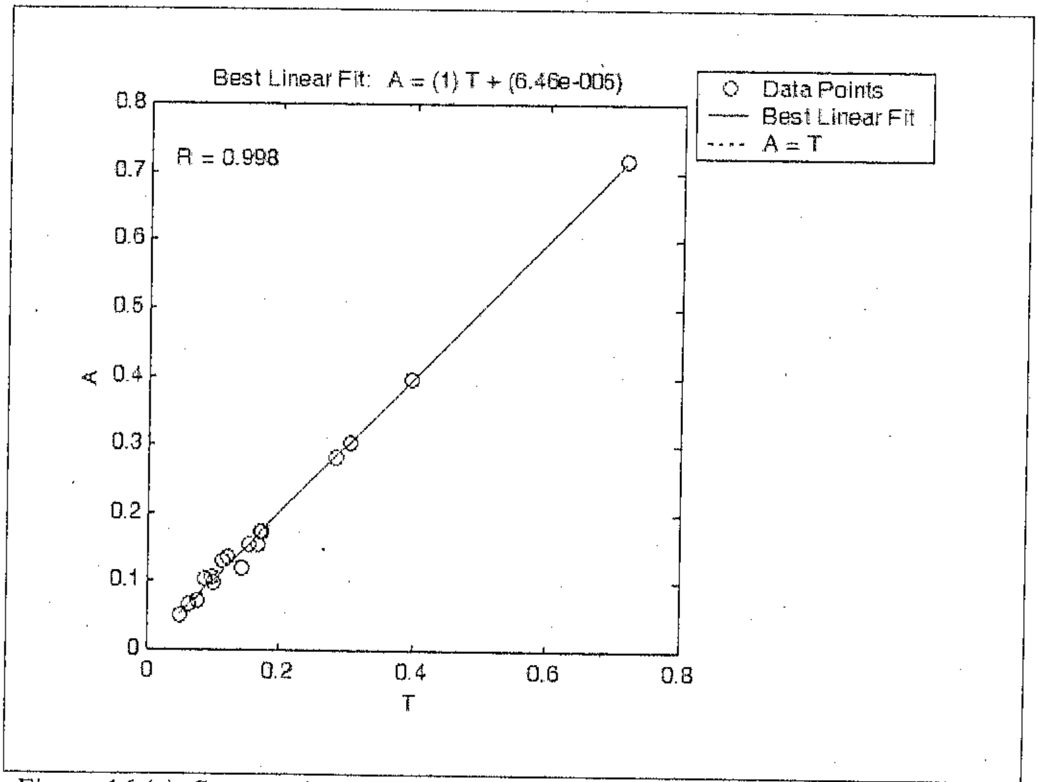

Figure 16 (a). Scatter plot for training data (17 points) for FFBP

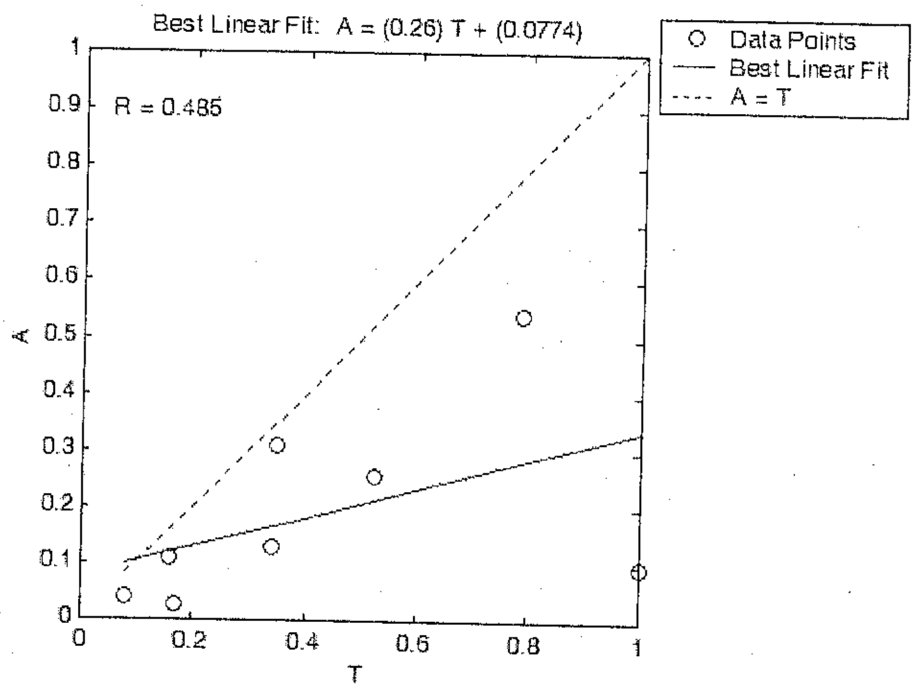

Figure 16. (b). Scatter plot for testing data (8 points) for FFBP

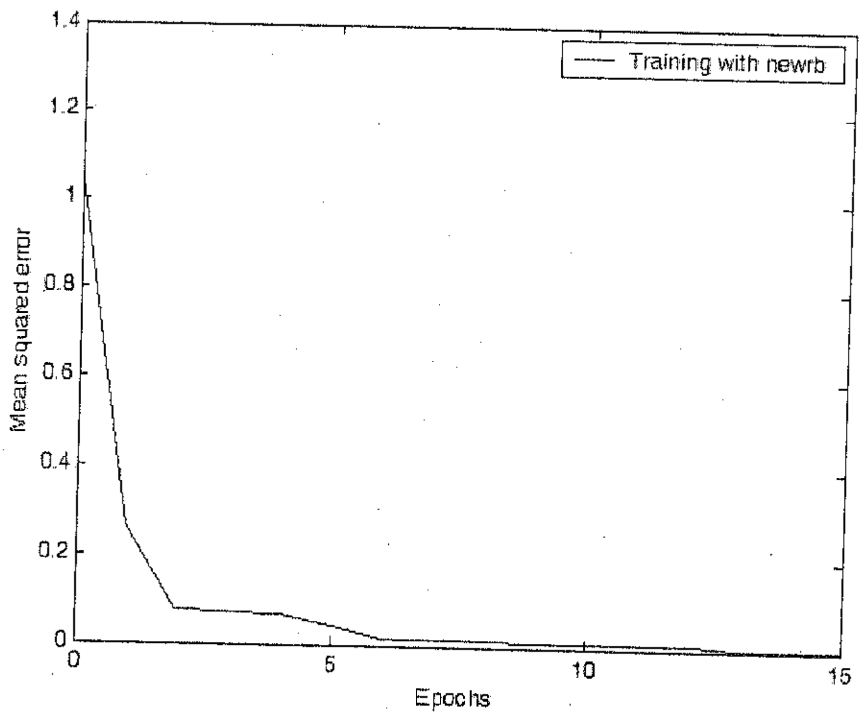

Figure 17. Epochs versus squared error values for RBF

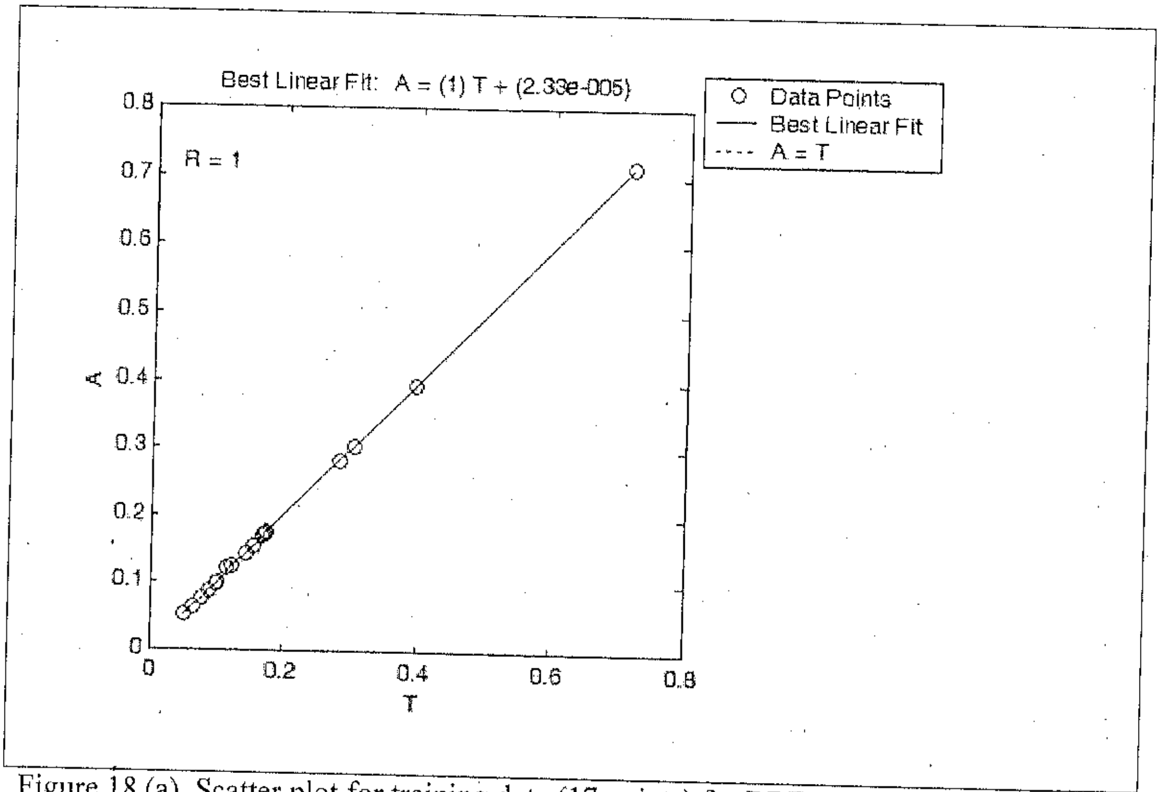

Figure 18 (a). Scatter plot for training data (17 points) for RBF

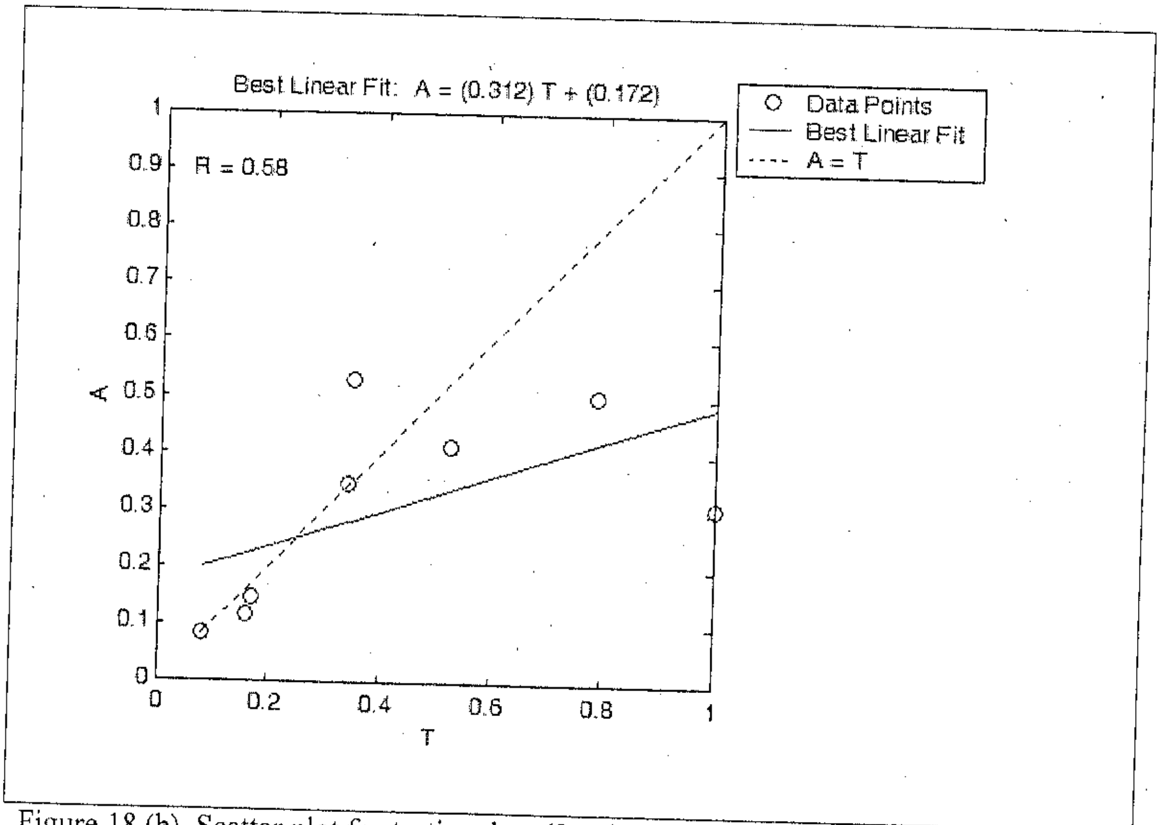

Figure 18 (b). Scatter plot for testing data (8 points) for RBF

Annexure-1: Literature Review

Prediction and Classification techniques

ASCE task committee (2000a, b) report discussed in detail about various ANN algorithms such as Feed Forward with Back Propagation (FFBP), Radial Basis Function (RBF), Generalized Regression Neural Networks (GRNN) in prediction category; Self Organizing Feature Map (SOFM) etc in classification category. Advantages and disadvantages of ANN along with various salient features were also discussed. Various applications of ANN in ground water, surface water, rainfall-runoff modeling etc were explored. Nagesh kumar (2004) discussed various perspectives of ANN. Sudheer and his team applied ANN in various perspectives in water resources (Sudheer 2005; Sudheer et al. 2006; Sudheer et al. 2008). Similarly Morse (1980), Jain and Dubes (1988) stressed the role of clustering and discussed various relevant algorithms. Ross (1995), Klir and Yuan (1995) and Zimmermann (1996) have given detailed description of Fuzzy Cluster Analysis.

Numerous authors used various types of prediction and classification techniques (ASCE 2000a,b). Agarwal and Singh (2004) developed multi layer Back Propagation Artificial Neural Network (BPANN) models to simulate rainfall-runoff process for two sub-basins of Narmada river, India considering three time scales with variable and uncertain data sets. Performance of BPANN models was compared with the developed Linear Transfer Function (LTF) model and was found superior. Altunkaynak (2007) developed back-propagation algorithm for predicting the temporal change in water levels of Lake Van in eastern Turkey.

Aqil et al. (2007) examined two types of neural network architectures i.e. feed forward and recurrent neural networks, and three types of training algorithm i.e. Levenberg–Marquardt, Bayesian regularization, and Gradient descent with momentum and adaptive learning rate backpropagation algorithms to predict flow for the case study of Cilalawi River in Indonesia. Six different neural network configurations were developed and examined in terms of optimum results for 1 to 5-h ahead prediction. Results of the study suggested that recurrent and feed forward network trained with Levenberg–Marquardt were able to forecast the catchment flow up to 5 h in advance with reasonable prediction accuracy.

Rai and Mathur (2008) developed back propagation feed forward Artificial Neural Networks (ANN) model for the computation of event-based temporal variation of sediment yield from watersheds. Training of the network was performed by using the gradient descent algorithm with automated Bayesian regularization with different ANN structures. Relative performance of the ANN model was evaluated by comparing the results obtained from the linear transfer function model and it is observed that ANN performed better. Sohail et al. (2008) adopted BPANN technique and the multivariate Auto Regressive and Moving Average (ARMA) models for two small sub-basins in a mountainous catchment of Tono area Japan for predicting runoff. It was found that the accuracy of prediction decreased with the increase of the time period for prediction. It was also found that the predictions by BPANN models were better than those by multivariate ARMA models for intense rains having complex rainfall runoff relationships especially in summer. On the other hand, both the modelling techniques yielded almost similar results for smaller rains in winter. Kisi (2008) applied Feed Forward with Back Propagation (FFBP), Radial Basis Function (RBF), Generalized Regression Neural Networks (GRNN) for

Gerdelli Station on Canakdere River and Isakoy Station on Goksudere River, in the Eastern Black Sea region of Turkey. He compared the outcome with Multiple Linear Regression. It was concluded that GRNN gave better performance than the FFNN and RBF techniques in one month ahead streamflow forecasting. Aksoy and Dahamsheh (2009) applied FFBP, RBF and generalized regression type ANNs, Multiple Linear Regression (MLR) models on monthly total precipitation at three meteorological stations (Baqura, Amman and Safawi) in Jordan.

Jingyi and Hall (2004) applied geographical approach (Residuals method), Ward's cluster method, Fuzzy c-means method, and Kohonen neural network to 86 sites in the Gan River Basin of Jiangxi Province and the Ming River Basin of Fujian Province in the southeast of China to delineate homogeneous regions based on site characteristics. It was concluded that Kohonen methodology is the preferred approach. Rao and Srinivas (2006a) studied the applicability of three hybrid-clustering algorithms, which use partitional clustering procedure to identify groups of similar catchments of Indiana, USA. The hierarchical clustering algorithms used were single linkage, complete linkage and Ward's algorithms, while the partitional clustering algorithm used was K-means algorithm. They also employed four cluster validity indices to determine their effectiveness in identifying optimal partition provided by the clustering algorithms. Rao and Srinivas (2006b) applied Fuzzy Cluster Analysis (FCA) to the above case study. They discussed the effectiveness of several fuzzy cluster validation measures in determining optimal partition provided by the FCA. Similar studies were reported by Cunderlik and Ouarda (2006), and Lin and Chen (2006). Raju et al. (2006) applied Kohonen Neural Networks for the case study of Jayakwadi irrigation project, India. Raju and Nagesh Kumar (2007) applied Fuzzy Cluster Analysis, Cluster Analysis and Kohonen Neural Networks for the classification of meteorological stations in Indian context. Rao and Srinivas (2008) performed detailed studies of classification techniques with reference to water resources.

Morphometric Analysis

Rao and Srinivas (2008) stressed the role of morphological parameters in flood frequency analysis. They mentioned that "Shape parameters of catchment such as form factor, compactness coefficient, elongation ratio or circulatory ratio may also be used as attributes to form regions for flood frequency analysis". They also mentioned that there was no criteria by which the superiority of one method can be established. It was felt that domain knowledge may be useful in choosing an appropriate algorithm.

Srinivasa et al. (2004) have used remote sensing and GIS techniques in morphometric analysis of sub-watersheds in Pawagada area of Tumkur district, Karnataka. Mishra et al. (2007) used average estimates of sediment yield from different sub-watersheds to prioritize check dam construction. Morphometric analysis has been carried out by Chopra et al. (2005) in Bhagra-Phungotri and Hara Maja sub-watersheds which drain into perennial Chaki Khad in Gurdaspur district, Punjab. Biswas et al (2002) performed similar analysis for a case study in Midnapore district of West Bengal, India. Machiwal et al. (2004) studied cost effective water harvesting structures for efficient utilization of water resources of Rajasthan, India. Rao and Kumar (2004) developed Spatial Decision Support System (SDSS) and applied to Tones watershed in India to compute soil loss, to prioritize watersheds, and to suggest various watershed management practices. Sewilam et al. (2007) developed rule-based decision support system (DSS) to investigate the feasibility of different morphological rehabilitation measures, to predict the impact on the morphological structure and to prepare a programme of morphological measures including cost estimation for watercourse in the German state of North Rhine-Westphalia (NRW). Singh et al. (2009) analyzed 13 dimensionless parameters namely, average slope of the watershed, Relief ratio, Relative relief, Main stream channel slope, Elongation ratio, Basin shape factor, Length–width ratio, Stream length ratio, Bifurcation ratio, Hypsometric analysis, Circulatory ratio, Ruggedness number, Drainage factor for 16 watersheds of the Chambal catchment, Rajasthan, India. They applied principal component analysis for screening out the parameters of least significance. It also helped to regroup the remaining variables into physically significant factors.

Annexure- 2: References

Agarwal A, Singh RD (2004) Runoff modelling through back propagation artificial neural network with variable rainfall-runoff data. Water Resour Manag 18: 285-300

Aksoy H, Dahamsheh A (2009) Artificial neural network models for forecasting monthly precipitation in Jordan. Stochastic Environme Resear and Risk Assess DOI. 10.1007/s00477-008-0267-x

Altunkaynak A (2007) Forecasting surface water level fluctuations of lake van by artificial neural networks, Water Resour Manag 21: 399-408

Aqil M, Kita I, Yano A, Nishiyama S (2007) Neural Networks for real time catchment flow modeling and prediction. Water Resour Manag 21: 1781-1796

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000a) Artificial neural networks in hydrology. I: preliminary concepts. J Hydrol Engineer ASCE 5: 115-123

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology. 2000b. Artificial neural networks in hydrology. II: hydrologic applications. J Hydrol Engineer ASCE 5: 124-137

Biswas S, Sudhakar S, Desai VR (2002) Remote sensing and geographic information system based approach for watershed conservation. J Survey Eng ASCE 128: 108-124

Chopra R, Dhiman RD, Sharma PK (2005) Morphometric analysis of sub-watersheds in gurdaspur district, punjab using remote sensing and GIS techniques. J Indian Soc of Remote Sens 33 : 531-539.

Cunderlik JM, Ouarda TBMJ (2006) Regional flood-duration-frequency modeling in the changing environment. J Hydrol 318: 276-291

Davies DL, Bouldin DW (1979) A cluster separation measure. IEEE Trans of Patte Analys and Mach Intell 1 : 224-227

Dunn JC (1974) Well separated clusters and optimal fuzzy partitions. J Cybern 4: 95-104

Garde RG (2006) River morphology, New Age International, New Delhi.

Ground Water Atlas of Rajasthan (2000) State Remote Sensing Application Center, Department of Science and Technology, Government of Rajasthan, Jaipur

Horton RE (1945) Erosional development of streams and their drainage basins: hydrological approach to quantitative morphology. Geolo Soc of Amer Bull 56: 275-370.

Jain AK, Dubes RC (1988) Algorithms for clustering cata. Prentice- Hall, Englewood Cliffs, New Jersey

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

Kaijun W (2008) Cluster Validation Toolbox, MATLAB central, file exchange, Mathworks http:// www.mathworks.com/matlabcentral/fileexchange/14620, Accessed September 2008

Kartalopoulos SV (2002) Understanding neural networks and fuzzy logic: basic concepts and applications. Prentice-Hall of India Private limited, New Delhi

Kisi O (2008) River flow forecasting and estimation using different artificial neural network techniques. Hydro Res 39.1: 27-40

Klir GJ, Yuan B (1995) Fuzzy sets and fuzzy logic: theory and applications, Prentice-Hall of India Private limited, New Delhi

Kohonen T (1989) Self organization and associative memory. Springer- Verlag, Berlin

Kothyari UC, Garde RJ (1991) Annual runoff estimation for catchments in India, J Water Resour Plan Manage ASCE 117 : 1-10

Kothyari UC (2006) Estimation of hydrologic variables from ungauged catchments using artificial neural networks. In: Predictions in Ungauged Basins for Sustainable Water Resources Planning and Management, Edited by K. Srinivasa Raju, Jain Brothers, New Delhi : 41-51

Lin GF, Chen LH (2006) Identification of homogeneous regions for regional frequency analysis using the self-organizing map. J Hydrol 324: 1-9

Liong SY, Tariq AAF, Lee KS (2004) Application of evolutionary algorithm in reservoir operations. J Insti of Eng (Singapore) 44: 39-54

Machiwal D, Jha MK, Singh PK, Mahnot SC, Gupta A (2004) Planning and design of cost-effective water harvesting structures for efficient utilization of scarce water resources in semi-arid regions of rajasthan, India. Water Resour Manag 18: 219–235

Mishra A, Kar S, Singh VP (2007) Prioritizing structural management by quantifying the effect of land use and land cover on watershed runoff and sediment yield. Water Resour Manag 21: 1899-1913

Morse JN (1980) Reducing the size of the nondominated set; pruning by clustering. Compu and Oper Res 7: 55-66

Nagesh Kumar D (2004) ANN applications in hydrology - merits and demerits, In: Integrated Water Resources Planning and Management, Edited by K. Srinivasa Raju, A.K.Sarkar and Motilal Dash, Jain Brothers, New Delhi: 31-42

Nantiyal MD (1994) Morphometric analysis of a drainage basin, district Dehradun, Uttar Pradesh. J Indian Soc of Remote Sens 22: 251-261

Pomerol JCh, Romero SB (2000) Multicriterion decision in management: principles and practice. Kluwer Academic, Netherlands

Rai RK, Mathur BS (2008) Event-based sediment yield modeling using artificial neural network. Water Resour Manag 22: 423-441

Raju KS, Kumar DN, Duckstein L (2006) Neural networks and multicriterion analysis for sustainable irrigation Planning. Compu and Oper Res 33: 1138-1153

Raju KS, Nagesh Kumar, D (2007) Classification of indian meteorological stations using cluster and fuzzy cluster analysis, and Kohonen artificial neural networks. Nordic Hydrology 38: 303–314

Rao AR, Srinivas VV (2006a) Regionalization of watersheds by hybrid-cluster analysis. J Hydrol 318: 37-56

Rao AR, Srinivas VV (2006b) Regionalization of watersheds by fuzzy cluster analysis. J Hydrol 318: 57-79

Rao AR, Srinivas VV (2008) Regionalization of watersheds an approach based on cluster analysis, 58, Water Science and Technology Library, Springer

Rao KHVD, Kumar DS (2004) Spatial decision support system for watershed management. Water Resour Manag 18: 407-423

Ratnam KN, Srivastava YK, Rao VV, Amminedu E, Murthy KSR (2005) Check dam positioning by prioritization of micro-watersheds using SYI model and morphometric analysis- remote sensing and GIS perspective. J Indian Soc of Remote Sens 33: 25-38

Ross TJ (1995) Fuzzy logic with engineering applications. McGraw-Hill, New York

Rousseeuw PJ (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J of Compu and Appl Mathe 20: 53–65

Rumelhart DE, Hinton GE, Williams RJ (1986) Learning internal representation by back-propagating errors. Nature 32: 533-536

Schumn SA (1956) Evolution of drainage systems and slopes in badlands at Perth Amboy, New Jersey. Geolo Soc of Amer Bull 67 : 597-646

Sewilam H, Bartusseck S, Nacken H (2007) Rule-based decision support system for the morphological rehabilitation of watercourses. Water Resour Manag 21: 2037-2047

Sharma V, Negi SC, Rudra RP, Yang S (2003) Neural networks for predicting nitrate-nitrogen in drainage water. Agri Wat Manag 63: 169-183

Singh S, Singh MC (1997) Morphometric analysis of Kanhar river basin. Nati Geogra Jour of India 43: 31-43

Singh PK, Kumar V, Purohit RC, Kothari M, Dashora PK (2009) Application of principal component analysis in grouping geomorphic parameters for hydrologic modeling. Water Resour Manag 23: 325-339

Sohail A, Watanabe K, Takeuchi S (2008) Runoff analysis for a small watershed of tono area japan by back propagation artificial neural network with seasonal data. Water Resour Manag 22: 1-22

Srinivasa VS, Govindaiah S, Honne Gowda H (2004) Morphometric analysis of sub-watersheds in the pawagada area of tumkur district South India using remote sensing and GIS techniques. J Indian Soc of Remote Sens 32: 351-362

State Water policy, Government of Rajasthan (1999). Department of Irrigation, (http://waterresources.rajasthan.gov.in/) Accessed on 2.7.09

Strahler AN (1957) Quantitative analysis of watershed geomorphology. Trans of Amer Geophy Uni, 38: 913-920

Sudheer KP, Jain SK (2003) Radial basis function neural network for modeling rating curves. J Hydrol Engineer ASCE 8: 161-164

Sudheer KP (2005) Knowledge extraction from trained neural network river flow models. J Hydrol Engineer ASCE 10: 264-269

Sudheer KP, Chaubey I, Garg V (2006) Lake water quality assessment from landsat TM data using neural network: an approach to optimal band combination selection. J Amer Wat Resour Soc 42: 1683-1695

Sudheer KP, Srinivasan K, Neelakanthan TR, Srinivas VV (2008) A nonlinear data-driven model for synthetic generation of annual streamflows. Hydrol Proce 22: 1831-1845

Watershed Atlas of Rajasthan (2000), State Remote Sensing Application Center, Department of Science and Technology, Government of Rajasthan, Jaipur

Zhang, GP (2007), Avoiding pitfalls in neural network research. IEEE Trans.Syst. Man, Cybern. C, Appl. Rev., 37: 3–16.

Zhang ZX, Kushwaha RL (1999) Applications of neural networks to simulate soil–tool interaction and soil behavior. Cana Agr Eng 41: 119–125

Zimmermann HJ (1996) Fuzzy set theory and its applications. Allied Publishers Limited, New Delhi

www.mathworks.com

Annexure-3: Acknowledgements

Principal investigator acknowledge the help provided by Indian National Committee on Hydrology (INCOH), Ministry of Water Resources (MOWR), Government of India for sanctioning the project entitled "Artificial Neural Networks for Water Resources Planning: An Innovative Approach".

I am grateful to experts present during INCOH R & D sessions held at Varanasi, Uttarpradesh; Udaipur, Rajasthan; Chennai, Tamilnadu who have given valuable advices about the project which allowed me to think beyond Artificial Neural Networks for refining the objectives. This paved the way for fuzzy cluster analysis, cluster validation indices, analyzation of the problem in morphometric analysis both in classification and prediction environment.

I am grateful to Dr R.D.Singh Director, National Institute of Hydrology, Roorkee; Dr Rakesh Kumar, Member Secretary, INCOH; Mr Hussain, Director, R & D, Ministry of Water Resources, Government of India; Dr K.D.Sharma, former Director, National Institute of Hydrology, Roorkee; Dr R.Jha, former member secretary, INCOH; Dr K.K.S.Bhatia, Scientist F (retired), National Institute of Hydrology, Roorkee for giving encouragement and valuable suggestions from time to time.

I express my deep gratitude to Dr N.K.Kalra, Project Director, State Remote Sensing Application Center (SRSAC), Jodhpur for providing the necessary data and maps. I benefited a lot from him in the form of discussions, suggestions and enjoyed the hospitality provided by him. I am thankful to Dr N Bhatia, Scientist, SRSAC, who accompanied and helped me in the field visits and discussions. Due to him only I am able to see the Kherthal watershed at microlevel in all the perspectives and interacted with relevant stake holders. I am grateful to all the staff at SRSAC for their joyful support during my long stays at Jodhpur and during field visits. I am very much benefited by the discussion with stake holders including the farmers, officials who have given me multiple perspectives about the watershed.

I am thankful to Prof D.Nagesh Kumar, Civil Engineering Department, Indian Institute of Science, Bangalore for providing valuable suggestions, moral support and relevant study material. I benefited a lot with publications of Dr V.V.Srinivas, Civil Engineering Department, Indian Institute of Science, Bangalore and Dr K.P.Sudheer, Civil Engineering Department, Indian Institute of Technology, Chennai. I am thankful to them.

I am grateful to reviewers for their valuable suggestions. I am thankful to Prof V.S.Rao, Director, BITS- Pilani, Hyderabad campus for providing all the facilities, suggestions for successful completion of the project. I am also thankful to Prof A.K.Sarkar, Dean, Instruction Division, BITS-Pilani for providing support from time to time.

K.Srinivasa Raju