
Volume 5 Issue 2 June 2011
ISSN 1570-6443
ELSEVIER
Journal of
Hydro-environment
Research
Journal of the Asia Pacific Division of the International
Association of Hydro-environment Engineering and Research
IAHR
AIRH
ScienceDirect

Available online at www.sciencedirect.com
ScienceDirect

Journal of Hydro-environment Research 5 (2011) 101–109

Journal of Hydro-environment Research

www.elsevier.com/locate/jher

Research paper

Classification of microwatersheds based on morphological characteristics

K. Srinivasa Raju [a], D. Nagesh Kumar [b,*]

[a] Department of Civil Engineering, Birla Institute of Technology and Science – Pilani, Hyderabad Campus, Hyderabad, India
[b] Department of Civil Engineering, Indian Institute of Science, Bangalore 560012, India

Received 7 February 2010; revised 15 April 2010; accepted 10 September 2010

Abstract

Three classification techniques, namely, K-means Cluster Analysis (KCA), Fuzzy Cluster Analysis (FCA), and Kohonen Neural Networks (KNN) were employed to group 25 microwatersheds of Kherthal watershed, Rajasthan into homogeneous groups for formulating the basis for suitable conservation and management practices. 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 used for the classification. Optimal number of groups is chosen, based on two cluster validation indices Davies–Bouldin and Dunn's. Comparative analysis of various clustering techniques revealed that 13 microwatersheds out of 25 are commonly suggested by KCA, FCA and KNN i.e., 52%; 17 microwatersheds out of 25 i.e., 68% are commonly suggested by KCA and FCA whereas these are 16 out of 25 in FCA and KNN (64%) and 15 out of 25 in KNN and CA (60%). It is observed from KNN sensitivity analysis that effect of various number of epochs (1000, 3000, 5000) and learning rates (0.01, 0.1–0.9) on total squared error values is significant even though no fixed trend is observed. Sensitivity analysis studies revealed that microwatersheds have occupied all the groups even though their number in each group is different in case of further increase in the number of groups from 5 to 6, 7 and 8.

Keywords: Watershed; Classification; Cluster validation indices; Morphology

1. Introduction

Watershed can be considered as a unit/block for planning of water and soil conservation programmes. Significant factors that affect planning and development of watershed are soil erosion, land use, deposition of sediments and water resources. These factors which may not be in the suitable levels may result in deterioration of the watershed. On the other hand, geomorphometric analysis received wide attention and acceptance from hydrologists and geomorphologists due to its ability of analyzing the watershed for various complex physical processes and hydrologic behavior for possible improvements in general and in low flow/drought situation in specific. Keeping this in view, integration of geomorphological parameters with hydrological characteristics of the watersheds is essential which will enable better planning and formulating the appropriate strategies for suitable conservation and management practices. Strategies thus developed can be implemented in a prioritized manner, if necessary, due to involvement of huge investment. In this process, it is expected to improve/rehabilitate the watershed(s) in a systematic and sustainable way (State Water Policy, Government of Rajasthan, 1999).

In addition, each watershed may not require same conservative measures/treatment. So, instead of analyzing each watershed for its rehabilitation/improvement, watersheds that are similar can be grouped so that problem can be tackled in the form of groups instead of individually for possible improvements. Clustering algorithms can be explored in this regard to form the groups in a small fraction of the time as compared to manual grouping, particularly if many criteria are associated with

* Corresponding author. Tel.: +91 80 2293 2666; fax: +91 80 2360 0404.
E-mail address: nagesh@civil.iisc.ernet.in (D. Nagesh Kumar).

1570-6443/$ - see front matter
doi:10.1016/j.jher.2010.09.002

watersheds. Keeping this in view, clustering techniques for grouping and cluster validation indices for determining optimal number of groups for evaluation of geomorphological characteristics are employed and is the basis for formulating the objectives for the present study which are mentioned below:

1. Exploring the applicability of K-means Cluster Analysis (KCA), Fuzzy Cluster Analysis (FCA) and Kohonen Neural Networks (KNN) for classification of microwatersheds into homogeneous groups based on 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).
2. Exploring the applicability of cluster validation indices, namely, Davies–Bouldin and Dunn's to determine the optimal number of clusters/groups of microwatersheds.
3. Sensitivity analysis of selected parameters

In the present paper, case study of Kherhtal watershed, Pali district, Rajasthan, India is chosen for microwatersheds to validate the above methodology for formulating the basis for suitable conservation and management practices. Same methodology can be applied to larger watersheds (scale up approach) or smaller watersheds (scale down approach). Brief literature review is presented below.

2. Literature review

2.1. Morphometric analysis

Morphometric analysis provides quantitative description of the basin geometry to understand mainly geological and geometric history of drainage basin (Strahler, 1957). The geomorphological properties that are important from the hydrological studies point of view include the linear, aerial and relief aspect of the watersheds i.e., catchment area, basin length, stream slope, mean basin elevation, rainfall, drainage density, bifurcation ratio, stream frequency, length of overland flow, form factor, shape factor, elongation ratio, circulatory ratio, compactness coefficient and texture ratio (Nautiyal, 1994). Rao and Srinivas (2008) also stressed the role of morphological parameters.

Biswas et al. (2002) performed morphometric analysis with parameters such as bifurcation ratio, drainage density, stream frequency, texture ratio, form factor, circularity ratio, elongation ratio for a case study of a watershed in Midnapore District of West Bengal, India. Rao and Kumar (2004) developed Spatial Decision Support System (SDSS) and applied to Tones watershed in India to compute soil loss, prioritize watersheds and to suggest various watershed management practices. Similar studies are reported by Srinivasa et al. (2004), Chopra et al. (2005), Ratnam et al. (2005), Sewilam et al. (2007). Garde (2006) studied river morphology in various perspectives. Mishra et al. (2007) used average estimates of sediment yield from different subwatersheds to prioritize check dam construction. 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 and drainage factor for 16 watersheds of Chambal catchment, Rajasthan, India. They applied principal component analysis for screening out the parameters of least significance. It was concluded that the study helped to regroup the remaining variables into physically significant factors.

2.2. Classification techniques

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.

ASCE Task Committee (2000a,b) report discussed self organizing feature maps used for classification purpose, their salient features and applications. Anand Raj and Nagesh Kumar (1998) presented an approach for ranking multi-criterion river basin planning alternatives using fuzzy numbers. Rao and Srinivas (2008) explained various classification algorithms related to K-means, fuzzy c-means and self organizing feature maps and their application to various case studies. Jain and Dubes (1988) stressed the role of clustering and discussed various relevant algorithms. Ross (1995) has given detailed description of Fuzzy Cluster Analysis.

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. Kothyari (2006) estimated hydrological variables from ungauged catchments using Artificial Neural Networks (ANN). Similar studies were reported by Rao and Srinivas (2006a,b), Lin and Chen (2006), Rao and Srinivas (2008). Raju and Nagesh Kumar (2007) applied Fuzzy Cluster Analysis, K-means Cluster Analysis and Kohonen Neural Networks for classification of meteorological stations in India (Raju and Nagesh Kumar, 2010).

It is observed from the above literature review that no study is reported in which

1. Clustering techniques and morphological parameters are integrated for a real world case study at microlevel of watersheds.
2. Methodology for determination of optimal/ideal number groups for effective conservation purposes in morphological perspective.

Adequately addressing the above lacunae is the basis for formulating the objectives as mentioned at the end of introduction.

3. Classification techniques employed

3.1. Classification methodology

In the present study, practical applicability of three classification techniques, namely, K-means Cluster Analysis (KCA), Fuzzy Cluster Analysis (FCA), and Kohonen Neural Networks (KNN) is explored for grouping 25 microwatersheds of Kherthal watershed, Pali District, Rajasthan. These techniques are explained briefly below.

K-means Cluster Analysis (KCA) partitions data sets into relatively homogeneous groups and used to minimize intra-cluster sums of squares of differences to obtain the final classification (Jain and Dubes, 1988). In KCA, each cluster is represented by its mean of feature vectors within the cluster (Rao and Srinivas, 2008). In this technique, data sets are grouped so that each data set is assigned to one group of the K (fixed number) groups. The sum of the squared differences of each criterion from its assigned cluster mean is used as the objective function. More information is available in Raju and Nagesh Kumar (2010). Important input parameters for KCA are number of epochs and tolerance criterion.

In Fuzzy Cluster Analysis each data set 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 set to a cluster center, weighted by that data set's membership grade. In other words, objective is to represent the similarity which the data set shares with each cluster having a membership function, whose value lies between zero and one. Each data set has a membership in every cluster (Ross, 1995) but degree of membership varies from cluster to cluster (between zero and one). The sum of the membership values for each data set will be equal to 1. Various steps in Fuzzy Cluster Analysis are available in Ross (1995), Jingyi and Hall (2004), Rao and Srinivas (2006a), Rao and Srinivas (2008). Important input parameters for FCA are number of iterations and tolerance criterion.

Kohonen Neural Networks (KNN) are 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 are to be partitioned. Each neuron in the output is interconnected with each of those in the input layer by a set of weights or a weight vector, e.g., the jth output neuron has a weight vector connecting to input neurons, $w_j = \{w_{ji}\}$, $i = 1,2,\ldots,M$. 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 facilitate assessing their similarity with others (Kohonen, 1989). Important input parameters for KNN are learning rate, conscience rate, and number of epochs and tolerance criterion. Fig. 1 presents schematic diagram of Kohonen Neural Networks

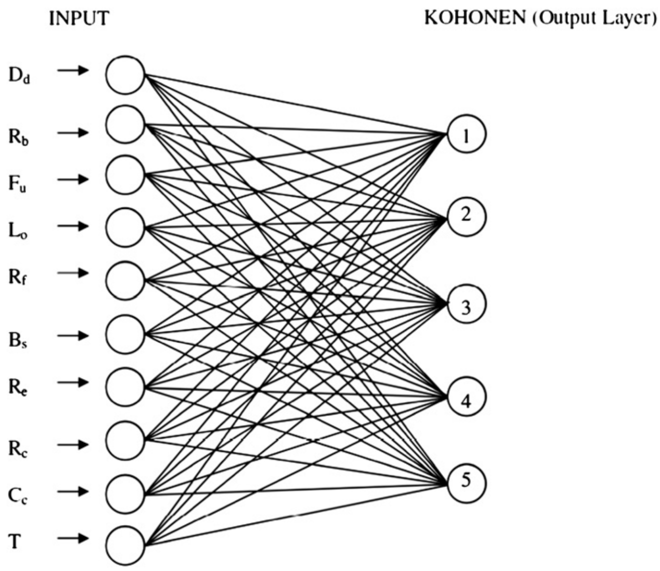

Fig. 1. Schematic diagram of Kohonen neural networks.

relevant to the planning problem where M and N are 10 and 5 respectively.

3.2. Cluster validation indices

Cluster validation indices are used to determine the optimal number of clusters in a data set. These are computed based on the outcome of clustering algorithms. Two cluster validation indices (Satyanarayana and Srinivas, 2008), namely, Davies–Bouldin Index and Dunn's Index are explored to determine the optimal number of clusters/groups which are explained as follows:

Davies–Bouldin Index (Davies and Bouldin, 1979; Jain and Dubes, 1988; Rao and Srinivas, 2008; Raju and Nagesh Kumar, 2010) is a function of the ratio of the sum of intra-cluster scatter to inter-cluster separation and is given by

$$DB(U) = \frac{1}{K}\sum_{i=1}^{K} \max\left[\frac{\Delta(X_i)+\Delta(X_j)}{\delta(X_i,X_j)}\right], \quad i \neq j \tag{1}$$

where $\delta(X_i, X_j)$ defines the inter-cluster distance between clusters X_i and X_j; $\Delta(X_i)$ represents intra-cluster distance of cluster X_i and K is the number of clusters of partition U. In this case, small index value represents good clusters, i.e., clusters are compact and their centers are far away from each other. The cluster configuration that minimizes $DB(U)$ is taken as the optimal number of clusters K.

Dunn's Index (Dunn, 1974; Rao and Srinivas, 2008; Raju and Nagesh Kumar, 2010) is defined as:

$$D(U) = \min_{1\le i\le K}\left\{\min_{\substack{1\le j\le K\\ j\neq i}}\left[\frac{\delta(X_i,X_j)}{\max\limits_{1\le k\le K}(\Delta(X_K))}\right]\right\} \tag{2}$$

where $\delta(X_i, X_j)$ and $\Delta(X_K)$ are defined as above. The main goal of this measure is to maximize inter-cluster distances while minimizing intra-cluster distances. The number of clusters that maximizes $D(U)$ is taken as the optimal number of clusters K.

4. Case study and analysis of data

Kherthal watershed, Pali District, Rajasthan, India is considered as a case study, which is located between 24°51′ to 24°58′ North latitudes and 73°8′ to 73°19′ East longitudes. Area of the watershed is 158.93 km^2 (Ground Water Atlas of Rajasthan, 2000; Watershed Atlas of Rajasthan, 2000). Some of the data are inferred from IRS-LISS-III imageries of the case study area. In addition, Survey Of India (SOI) topo sheets 45 H/1 and 45 H/5 on a scale of 1:50,000 are also used.

The area is in a 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 month while May and June are the hottest months. Normal annual 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 (winter i.e., November–February) season are wheat and mustard whereas bajra, pulses etc. are grown in Kharif (summer i.e., June–October) season. Groundwater is the major source for irrigation and groundwater potential in the region is 40–80 m^3/day. There are 15–18 wells maintained by the Government. Quality of groundwater is significantly influenced by semi-arid climate and hydrogeological diversity. Salinity, sodicity and fluoride content are the major factors effecting the groundwater quality. Quality of water varies from potable to un-potable with electrical conductivity (E_c) less than 2000 μS/cm. Nitrate content ranges from 50 to 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 denudated hills, pediments, buried pediments and valley fields (Ground Water Atlas of Rajasthan, 2000; Watershed Atlas of Rajasthan, 2000).

Ten morphological 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) and texture ratio (T) are chosen for classification/grouping purpose and based on the studies reported by Biswas et al. (2002), Chopra et al. (2005), Ratnam et al. (2005), Garde (2006) and extensive discussion with experts. Description of various geomorphological parameters is available in Garde (2006).

In the present study, catchment area, basin length, channel slope, land use, mean basin elevation, are measured/estimated/inferred from various data sources. In addition, ten already mentioned parameters are computed for all the 25 microwatersheds. Mathematical expressions for deriving the above parameters are presented in Table 1 along with corresponding values for microwatershed number 3. Table 2 presents the area of microwatersheds and the payoff matrix consisting of 25 microwatersheds and 10 parameters. These 10 parameters are considered as the classification criteria for grouping microwatersheds. It is observed from Table 2 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, 22.250), 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).

The estimated parameters are normalized based on the methodology suggested by Biswas et al. (2002), Ratnam et al. (2005) and discussion with experts. Normalized values of criterion j for microwatershed i is defined as

$$y_{ij} = \frac{x_{ij}}{\left(x_{j\text{ideal}}\right)} \tag{3}$$

where x_{ij} is jth criterion for the ith microwatershed, $x_{j\text{ideal}}$ is ideal/desirable value of jth criterion among the 25 microwatersheds.

For example, in the case of drainage density, the maximum value is the ideal value i.e., 7.985 whereas in the case of compactness coefficient, the minimum value is ideal being desirable and is 1.228 (Ratnam et al., 2005). Table 2 presents ideal values (maximum or minimum depending on the parameters) at the last row. In case of microwatershed 1, normalized value for drainage density is 0.384/7.985 = 0.048 and the compactness coefficient is 1.321/1.228 = 1.076. Other values are similarly normalized and presented in Table 3. Even though, form factor, shape factor, bifurcation ratio, elongation ratio, circulatory ratio and compactness coefficient do not have any units, these are also normalized because of the differences in their variance, relative magnitude etc (Rao and Srinivas,

Table 1
Morphometric and derived parameters.

Parameters	Unit	Formula	Derived values for microwatershed number 3
Basin length	km	$L_b = 1.312A^{0.568}$	5.508
Drainage density	km^{-1}	$D_d = L/A$	3.834
Bifurcation ratio	No unit	$R_b = N_u/N_{u+1}$	2.292,1,800; Average is 11.292/3 = 3.764
Stream frequency	km^{-2}	$F_u = N/A$	8.480
Length of overland flow	km	$L_o = 0.5/D_d$	0.130
Form factor	No unit	$R_f = A/L_b^2$	0.412
Shape factor	No unit	$B_s = L_b^2/A$	2.427
Elongation ratio	No unit	$R_e = 1.128A^{0.5}/L_b$	0.724
Circulatory ratio	No unit	$R_c = 12.57A/P^2$	0.539
Compactness coefficient	No unit	$C_c = 0.2821P/A^{0.5}$	1.361
Texture ratio	km^{-1}	$T = N_1/P$	3.222

where for microwatershed number 3, A = Area of the microwatershed (km^2) = 12.500; P = Perimeter (km) = 17.067; L = Total length of all streams of all orders (km) = 47.933; 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 2
Areas of microwatersheds and morphological parameters of microwatersheds in Kherthal watershed.

Microwatershed No.	Parameter										
	Area (km^2)	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	2.750	0.384	0.000	0.364	1.303	0.506	1.975	0.803	0.573	1.321	0.129
2	50.500	3.392	1.701	6.337	0.147	0.341	2.934	0.658	0.422	1.539	4.152
3	12.500	3.834	3.764	8.480	0.130	0.412	2.427	0.724	0.539	1.362	3.223
4	6.750	5.126	1.898	15.556	0.098	0.448	2.232	0.755	0.490	1.429	4.330
5	3.000	4.577	1.839	14.667	0.109	0.500	1.999	0.798	0.587	1.306	3.118
6	4.250	2.696	3.060	9.647	0.185	0.477	2.096	0.779	0.621	1.269	2.587
7	4.000	7.985	2.472	22.250	0.063	0.481	2.078	0.782	0.350	1.692	4.086
8	0.500	2.226	0.000	4.000	0.225	0.638	1.566	0.901	0.664	1.228	0.650
9	2.000	4.477	2.938	12.500	0.112	0.529	1.892	0.820	0.593	1.299	2.304
10	0.750	4.436	2.750	13.333	0.113	0.604	1.655	0.877	0.500	1.415	1.612
11	1.750	3.969	2.200	9.143	0.126	0.538	1.857	0.828	0.185	2.323	1.010
12	10.500	4.700	1.993	13.238	0.106	0.422	2.370	0.733	0.274	1.910	3.783
13	3.200	3.974	2.292	10.625	0.126	0.496	2.016	0.794	0.347	1.697	1.952
14	1.300	3.464	2.667	8.462	0.144	0.561	1.784	0.845	0.432	1.522	1.300
15	2.500	4.464	1.843	11.200	0.112	0.513	1.950	0.808	0.651	1.240	2.303
16	1.500	2.798	2.333	6.667	0.179	0.550	1.819	0.836	0.429	1.527	1.056
17	4.250	4.303	1.720	9.176	0.116	0.477	2.096	0.779	0.414	1.555	1.760
18	14.000	4.071	2.224	8.929	0.123	0.406	2.465	0.719	0.482	1.441	3.507
19	2.500	4.870	2.150	12.800	0.103	0.513	1.950	0.808	0.587	1.306	2.460
20	6.750	3.520	1.775	6.074	0.142	0.448	2.232	0.755	0.336	1.726	1.447
21	2.250	3.385	1.000	5.333	0.148	0.520	1.922	0.814	0.554	1.344	0.840
22	1.130	1.607	0.000	1.770	0.311	0.571	1.750	0.853	0.489	1.430	0.371
23	4.500	1.979	3.400	3.333	0.253	0.473	2.112	0.776	0.216	2.154	0.556
24	13.550	2.814	2.567	4.576	0.178	0.408	2.454	0.720	0.446	1.498	1.841
25	2.250	3.829	2.167	8.444	0.131	0.520	1.922	0.814	0.452	1.488	1.643
Nature of extreme		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 3
Normalized morphological parameters of microwatersheds of Kherthal watershed.

Microwatrshed No.	Parameter									
	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

2008). This also helped to maintain consistency and uniformity in the methodology. These normalized data are used for classifying the microwatersheds.

5. Results and discussion

5.1. Determination of optimal number of clusters based on KCA methodology

An effort was made to ascertain the optimal number of groups for the present problem. K-means 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 with K ranging from 3 to 6 groups (totalling to 4 in number) with 1000 iterations. The range is based on the past literature (Watershed Atlas of Rajasthan, 2000; Rao and Srinivas, 2008) and discussion with experts. 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 importance is assigned to all the 10 parameters i.e., contribution or impact of each parameter in the objective function is assumed to be same for the present classification problem.

Davies–Bouldin index (Davies and Bouldin, 1979) and Dunn's index (Dunn, 1974) available in Cluster Validity Analysis Platform (CVAP) developed by Kaijun (2008) is used in the present study to find the optimal number of groups. Davies–Bouldin indices based on K-means Cluster Analysis as the classification algorithm for 3, 4, 5 and 6 number of groups are 1.023, 0.926, 0.956, 0.841 (optimal i.e., low value is 0.841 for number of groups 6) whereas for Dunn's index these are 1.333, 1.670, 1.340, 1.378 (optimal i.e., high value is 1.670 for number of groups 4). It can be seen that both 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, 5 groups were chosen for classifying the microwatersheds (Watershed Atlas of Rajasthan, 2000). The chosen number of groups of 5 was adopted for further classification analysis.

5.2. Application of KCA

Table 4 presents classification of microwatersheds obtained based on KCA methodology with 5 number of groups. On adopting KCA methodology, the number of microwatersheds falling in groups 1–5 is 3, 6, 6, 6 and 4. The numerals in parenthesis in Table 4 represent total number of microwatersheds in that group. It is observed that 2 groups are having 3 and 4 microwatersheds and other groups are having 6 each. Similarly, effort is also made to ascertain the occupancy rate of microwatersheds with further increase in the number of groups to 6, 7 and 8. It is observed (results not shown) that microwatersheds have occupied all the groups (as in some situations microwatersheds may not fall at all in one or more groups making the groups redundant) even though their number in each group is different. In this context, performance of the present classification algorithm is satisfactory.

5.3. Application of FCA

Fuzzy Cluster Analysis (www.mathworks.com) is used to classify microwatersheds. The membership value in each group indicates the probability for the microwatershed to be included in that specific group (Rao and Srinivas, 2008). Membership values of the 25 microwatersheds under each of the 5 groups are presented in Table 5. 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 0.086, 0.271, 0.161, 0.203 and 0.280. The sum of these values should be equal to 1 (Ross, 1995). The representative group for this microwatershed is group number 5 (having the maximum membership value of 0.280). Similarly all other microwatersheds were analyzed and presented in the last column of Table 5. Number of microwatersheds falling in groups 1–5 is 2, 7, 5, 5 and 6 respectively. The minimum number of microwatersheds in group 1 is 2 whereas maximum is 7 for group 2. The microwatershed with the highest membership value in a group is the representative microwatershed for that group. Table 4 presents classification of microwatersheds obtained based on FCA methodology. Similarly, effort was also made to examine the occupancy rate of microwatersheds on increasing the number of groups to 6, 7 and 8. It was observed that microwatersheds have occupied all the groups even though their number in each group is different.

Table 4
Classification of microwatersheds by FCA, KNN and KCA.

Group	Technique			Common microwatersheds based on			
	FCA	KNN	KCA	All 3 methods	FCA and KNN	FCA and KCA	KNN and KCA
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
			Total	13	16	17	15

Table 5
Membership values of the microwatersheds under each group showing the representative group of each microwatershed.

Microwatershed No.	Group					Representative group and corresponding membership value
	1	2	3	4	5	
1	0.086	0.271	0.161	0.203	0.280	5(0.280)
2	0.062	0.092	0.215	0.497	0.134	4(0.497)
3	0.032	0.242	0.099	0.474	0.153	4(0.474)
4	0.029	0.144	0.116	0.582	0.129	4(0.582)
5	0.012	0.776	0.034	0.106	0.072	2(0.776)
6	0.017	0.720	0.044	0.124	0.095	2(0.720)
7	0.131	0.108	0.356	0.228	0.177	3(0.356)
8	0.057	0.436	0.112	0.161	0.235	2(0.436)
9	0.010	0.833	0.026	0.065	0.066	2(0.833)
10	0.036	0.230	0.117	0.140	0.478	5(0.478)
11	0.872	0.014	0.062	0.022	0.030	1(0.872)
12	0.235	0.058	0.440	0.147	0.121	3(0.440)
13	0.027	0.011	0.885	0.028	0.049	3(0.885)
14	0.024	0.043	0.116	0.063	0.754	5(0.754)
15	0.011	0.843	0.027	0.062	0.057	2(0.843)
16	0.021	0.034	0.100	0.051	0.794	5(0.794)
17	0.031	0.045	0.460	0.139	0.326	3(0.460)
18	0.002	0.007	0.008	0.974	0.009	4(0.974)
19	0.004	0.918	0.012	0.034	0.031	2(0.918)
20	0.084	0.026	0.722	0.068	0.100	3(0.722)
21	0.032	0.377	0.094	0.158	0.338	2(0.377)
22	0.068	0.183	0.166	0.157	0.426	5(0.426)
23	0.842	0.017	0.075	0.028	0.038	1(0.842)
24	0.043	0.086	0.223	0.391	0.256	4(0.391)
25	0.007	0.020	0.042	0.037	0.895	5(0.895)

5.4. Application of KNN

Kohonen Neural Networks (www.mathworks.com) methodology is used to classify microwatersheds. The number of nodes in the input layer is 10 as shown in Fig. 1, representing 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). The nodes in output layer are five (optimal number of groups). Equal importance is assigned to all the 10 parameters i.e., contribution or impact of each parameter in the objective function is assumed to be same for the present classification problem.

The parameters used for training the algorithm are the number of groups equal to 5, learning rate of 0.01, conscience rate of 0.001, number of epochs equal to 1000 and elapsed time of 300 s. Initially weight values (i.e., connection strength between output and input neurons) are assumed as 0.5. The algorithm terminates after reaching prespecified number of epochs or elapsed time whichever occurs earlier. Table 6 presents the values of weights obtained after such process. It is observed that final values of the weights for the first output neuron (i.e., group 1) with input neurons (10 in this case) obtained after 1000 epochs for chosen parameters are (0.436, 0.610, 0.346, 0.113, 1.488, 1.276, 1.218, 2.342, 1.238, 0.352). Values of weights for other groups 2, 3, 4 and 5 are presented in Table 6. It is also observed that final weights are nowhere matching with the initially assumed weight values of 0.5. It is observed from Table 6 that 15 weight values out of 50 are less than initial weight values of 0.5 whereas 35 are above. It is noted that final values of weights of drainage density vary from 0.297 to 0.601. These are (0.189, 0.637), (0.215, 0.545), (0.088,0.327), (1.222, 1.664), (1.134,1.554), (1.104, 1.289), (1.466, 3.274), (1.047, 1.598), (0.163, 0.883) respectively for bifurcation ratio, stream frequency, length of overland flow, form factor, shape factor, elongation ratio, circulatory ratio, compactness coefficient and texture ratio. A wide variation is observed between the lower and upper values of final weights in case of circulatory ratio.

Table 4 presents classification of microwatersheds obtained based on KNN methodology. All microwatersheds have been grouped into the targeted 5 groups. The number of microwatersheds falling into the 5 groups is 4, 5, 5, 6 and 5 respectively with total squared error value of 6.42. Detailed computation of total squared error calculation is as follows: Normalized values for 10 criterion (or criterion vector), namely, drainage density

Table 6
Final weights for Kohonen neural networks.

Group	Parameter									
	D_d	R_b	F_u	L_o	R_f	B_s	R_e	R_c	C_c	T
1	0.436	0.610	0.346	0.113	1.488	1.276	1.218	2.342	1.238	0.352
2	0.527	0.634	0.545	0.095	1.482	1.266	1.217	3.274	1.047	0.592
3	0.463	0.620	0.390	0.115	1.395	1.353	1.180	1.466	1.598	0.412
4	0.601	0.637	0.541	0.088	1.222	1.554	1.104	2.467	1.214	0.883
5	0.297	0.189	0.215	0.327	1.664	1.134	1.289	3.007	1.096	0.163

Note: Refer to Table 1 for the name and mathematical expression for each of the 10 parameters.

(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) corresponding to alternative 14 clustered in group 1 (as presented in Table 3) are (0.434, 0.708, 0.380, 0.111, 1.645, 1.139, 1.283, 2.329, 1.240, 0.300). Correspondingly, final weight vectors for the first output neuron (i.e., group 1) with input neurons (above 10 criteria) obtained for chosen parameters as presented in Table 6 are (0.436, 0.610, 0.346, 0.113, 1.488, 1.276, 1.218, 2.342, 1.238, 0.352). Then squared error value of criterion vector of alternative 14 that clustered in group 1 is $[(0.434-0.436)^2+(0.708-0.610)^2+(0.380-0.346)^2+(0.111-0.113)^2+(1.645-1.488)^2+(1.139-1.276)^2+(1.283-1.218)^2+(2.329-2.342)^2+(1.240-1.238)^2+(0.300-0.352)^2=0.062$. With this back ground squared error values for groups 1–5 are computed as follows: Squared error values of alternatives 14, 16, 17, 25 from group 1 (Table 4) are 0.062, 0.054, 0.067, 0.019. Thus, total squared error value for group 1 is summation of 0.062, 0.054, 0.067, 0.019 which is equal to **0.202**; Squared error values of alternatives 5, 6, 9, 15, 19 from group 2 (Table 4) are 0.065, 0.102, 0.041, 0.086, 0.025. Thus, total squared error value for group 2 is summation of 0.065, 0.102, 0.041, 0.086, 0.025 which is equal to **0.319**. Total squared error value for group 3 (alternatives in this group are 11, 12, 13, 20, 23) is summation of 0.407, 0.337, 0.234, 0.211, 0.387 which is equal to **1.576**; Total squared error value for group 4 (alternatives in this group are 2, 3, 4, 7, 18, 24) is summation of 0.337, 0.399, 0.119, 0.834, 0.056, 0.393 which is equal to **2.138**; Similarly total squared error value for group 5 (alternatives in this group are 1, 8, 10, 21, 22) is summation of 0.669, 0.469, 0.729, 0.100, 0.216 which is equal to **2.183**; Overall squared error value $= 0.202+0.319+1.576+2.138+2.183=\mathbf{6.418\sim 6.42}$. It is noticed that distribution of microwatersheds amongst the five groups is almost even.

In KNN, the learning rate plays a major role, for a given number of epochs. An effort is made to examine its impact by conducting sensitivity analysis. Table 7 presents occupancy of microwatersheds in each group and total squared error values for various epochs (1000, 3000, 5000) and learning rates (0.01, 0.1–0.9). It is observed from Table 7 that total squared error values for learning rate of 0.01–0.9 vary from 6.339 to 10.906, 6.037 to 8.135 and 6.052 to 14.376 for 1000, 3000 and 5000 epochs respectively. Similarly, effort is also made to ascertain the distribution of microwatersheds in case of increase in the number of groups to 6, 7 and 8. It was observed that microwatersheds have occupied all the groups. From the extensive sensitivity analysis, it is observed that the effect of various learning rates and number of epochs is significant on the total squared error value even though no fixed trend is observed.

Comparative analysis of various clustering techniques is also performed. It can be observed from Table 4 that 13 microwatersheds out of 25 are commonly suggested by KCA, FCA and KNN i.e., 52% with 5, 1, 5, 2 in groups 2–5 leaving group 1 empty. It is also observed that 17 microwatersheds out of 25 i.e., 68% are suggested by both KCA and FCA whereas these are 16 in the case of FCA and KNN (64%); and 15 in the case of KNN and KCA (60%). It is also felt from the results pattern, sensitivity analysis and inferences from the discussions with experts that K-means Cluster Analysis may be used as the basis for clustering microwatersheds due to its advantage of less parameter requirement.

Table 7
Distribution of microwatersheds in each of the 5 groups and corresponding total squared error values for various learning rates and epochs for KNN.

Learning rate	Epochs		
	1000	3000	5000
0.01	4,5,5,6,5 (6.420)	5,5,5,5,5 (6.254)	5,5,5,5,5 (6.254)
0.1	5,4,5,5,6 (6.409)	5,4,5,5,6 (6.407)	5,4,5,5,6 (6.323)
0.2	5,5,4,5,6 (6.339)	5,4,5,6,5 (6.426)	6,5,4,5,5 (6.886)
0.3	4,5,5,5,6 (6.945)	5,5,5,6,4 (6.716)	4,6,4,4,7 (6.052)
0.4	6,4,6,4,5 (7.371)	4,5,6,5,5 (6.037)	5,4,5,6,5 (6.651)
0.5	6,5,6,4,4 (7.012)	5,5,5,5,5 (7.301)	6,4,4,6,5 (8.868)
0.6	6,4,4,4,7 (8.579)	6,4,6,4,5 (7.003)	3,6,3,8,5 (7.346)
0.7	6,4,4,8,3 (7.870)	4,5,4,7,5 (7.961)	4,6,5,5,5 (7.394)
0.8	6,4,4,6,5 (8.534)	8,4,5,3,5 (7.335)	8,3,4,8,2 (10.060)
0.9	5,4,5,2,9 (10.906)	7,7,2,7,2 (8.135)	3,12,6,2,2 (14.376)

Based on the classification of microwatersheds based on morphological characteristics, it can be concluded that all the 25 microwatersheds in Kherthal watershed will not require similar conservation measures/treatment. For example based on the outcome of KCA (with reference to Table 4) microwatersheds 5, 6, 9, 10, 15 and 19 are falling in group 2 which means that these six microwatersheds are similar in nature. These six require similar conservation measures/treatment. This facilitates the choice of strategies appropriate for each group in the Kherthal watershed in respect of agricultural practices and paves the way for efficient utilization of resources. Five groups suggested by KCA can be analyzed for conservation/treatment measures and priority level for each group can be decided by the water resources planner if situation warrants. A report based on this study is communicated to government agencies to prioritize and adopt suitable conservation measures in the Kherthal watershed.

6. Summary and conclusions

Three clustering techniques, namely, K-means Cluster Analysis (KCA), Fuzzy Cluster Analysis (FCA) and Kohonen Neural Networks (KNN) were employed to group the 25 microwatersheds of Kherthal watershed into 5 groups that can be used to formulate the basis for initiating suitable soil and water conservation measures. Ten morphological 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) and texture ratio (T) were used as the classification criteria.

Choosing multiple clustering algorithms and cluster validation indices enhances the decision making ability to choose the right clustering algorithm and number of groups, as different algorithms work with different methodologies. This also gives an opportunity to explore the application potential of various techniques including their parameter requirement and robustness in the result pattern.

It is inferred from the results of the above studies and inferences from each perspective, that the outcome varied from algorithm to algorithm with the chosen set of parameters and data availability, but the methodology remains the same which is the main target of the present study.

The following conclusions are drawn from the study:

1. Classification methodologies based on morphometric analysis can be extended for wider applications of watershed development and management.
2. The KCA methodology can be used as the basis for clustering microwatersheds due to its advantage of less parameter requirement and its wider applicability and acceptability to various case studies.
3. Methodology of determining optimal number of groups based on Davies−Bouldin Index and Dunn's Index can be used as the basis that can be used in clustering algorithms.
4. Comparative analysis of various clustering techniques revealed that 13 microwatersheds out of 25 are commonly suggested by KCA, FCA and KNN.
5. It is observed from KNN sensitivity analysis that effect of various number of epochs (1000, 3000, 5000) and learning rates (0.01, 0.1−0.9) on total squared error values is significant even though no fixed trend is observed.

Fuzzy Cluster Analysis as well as Kohonen Neural Networks can be explored as the basis for clustering techniques in further studies, along with cost analysis and homogeneity measures. If more data is available, two phase approach may be explored where classification algorithms can be used to formulate group-wise relationships between parameters of interest to ensure regional homogeneity for various planning purposes.

Acknowledgements

This work is partially supported by Indian National Committee on Hydrology (INCOH), Ministry of Water Resources, Govt. of India, through Project # 23/51/2006-R&D/203-14. First author is grateful to Dr N.K. Kalra, Project Director, State Remote Sensing Application Center (SRSAC), Jodhpur, Rajasthan, India for providing the necessary data and maps and valuable suggestions from time to time. He is also thankful to Dr N. Bhatia, Scientist, SRSAC, who accompanied him and helped him in the field visits and discussions.

References

Anand Raj, P., Nagesh Kumar, D., 1998. Ranking multi-criterion river basin planning alternatives using fuzzy numbers. Fuzzy Sets Syst. 100, 89−99.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000a. Artificial neural networks in hydrology. I: preliminary concepts. J. Hydrol. Eng. 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. Eng. ASCE 5, 124−137.

Biswas, S., Sudhakar, S., Desai, V.R., 2002. Remote sensing and geographic information system based approach for watershed conservation. J. Survey Eng. ASCE 128, 108−124.

Chopra, R., Dhiman, R.D., Sharma, P.K., 2005. Morphometric analysis of sub-watersheds in Gurdaspur District, Punjab using remote sensing and GIS techniques. J. Indian Soc. Remote Sens. 33, 531−539.

Davies, D.L., Bouldin, D.W., 1979. A cluster separation measure. IEEE Trans Pattern Anal. Mach. Intell. 1, 224−227.

Dunn, J.C., 1974. Well separated clusters and optimal fuzzy partitions. J. Cybern. 4, 95−104.

Garde, R.G., 2006. River Morphology. New Age International, New Delhi, India.

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

Jain, A.K., Dubes, R.C., 1988. Algorithms for Clustering Data. Prentice-Hall, Englewood Cliffs, New Jersey, USA.

Jingyi, Z., Hall, M.J., 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.

Kohonen, T., 1989. Self Organization and Associative Memory. Springer-Verlag, Berlin, Germany.

Kothyari, U.C., 2006. Estimation of hydrologic variables from ungauged catchments using artificial neural networks. In: Srinivasa Raju, K. (Ed.), Predictions in Ungauged Basins for Sustainable Water Resources Planning and Management. Jain Brothers, New Delhi, India, pp. 41−51.

Lin, G.F., Chen, L.H., 2006. Identification of homogeneous regions for regional frequency analysis using the self-organizing map. J. Hydrol. 324, 1−9.

Mishra, A., Kar, S., Singh, V.P., 2007. Prioritizing structural management by quantifying the effect of land use and land cover on watershed runoff and sediment yield. Water Resour. Manage. 21, 1899−1913.

Nautiyal, M.D., 1994. Morphometric analysis of a drainage basin, District Dehradun, Uttar Pradesh. J. Indian Soc. Remote Sens. 22, 251−261.

Raju, K.S., Nagesh Kumar, D., 2007. Classification of Indian meteorological stations using cluster and fuzzy cluster analysis, and Kohonen artificial neural networks. Nordic Hydrol. 38, 303−314.

Raju, K.S., Nagesh Kumar, D., 2010. Multicriterion Analysis in Engineering and Management. Prentice Hall of India, New Delhi.

Rao, A.R., Srinivas, V.V., 2006a. Regionalization of watersheds by hybrid-cluster analysis. J. Hydrol. 318, 37−56.

Rao, A.R., Srinivas, V.V., 2006b. Regionalization of watersheds by fuzzy cluster analysis. J. Hydrol. 318, 57−79.

Rao, A.R., Srinivas, V.V., 2008. Regionalization of Watersheds: An Approach Based on Cluster Analysis, vol. 58. Springer. Water Science and Technology Library.

Rao, K.H.V.D., Kumar, D.S., 2004. Spatial decision support system for watershed management. Water Resour. Manage. 18, 407−423.

Ratnam, K.N., Srivastava, Y.K., Rao, V.V., Amminedu, E., Murthy, K.S.R., 2005. Check dam positioning by prioritization of micro-watersheds using SYI model and morphometric analysis − remote sensing and GIS perspective. J. Indian Soc. Remote Sens. 33, 25−38.

Ross, T.J., 1995. Fuzzy Logic with Engineering Applications. McGraw-Hill, New York, USA.

Satyanarayana, P., Srinivas, V.V., 2008. Regional frequency analysis of precipitation using large-scale atmospheric variables. J. Geophys. Res. − Atmospheres 113, D24110.

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

Singh, P.K., Kumar, V., Purohit, R.C., Kothari, M., Dashora, P.K., 2009. Application of principal component analysis in grouping geomorphic parameters for hydrologic modeling. Water Resour. Manage. 23, 325−339.

Srinivasa, V.S., 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. Remote Sens. 32, 351−362.

State Water Policy, Government of Rajasthan, 1999. Department of Irrigation. http://waterresources.rajasthan.gov.in/.

Strahler, A.N., 1957. Quantitative analysis of watershed geomorphology. Trans. Am. Geophys. Uni. 38, 913−920.

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