Water Resour Manage (2009) 23:325–339
DOI 10.1007/s11269-008-9277-1

Application of Principal Component Analysis in Grouping Geomorphic Parameters for Hydrologic Modeling

P. K. Singh · Virendra Kumar · R. C. Purohit · Mahesh Kothari · P. K. Dashora

Received: 24 November 2006 / Accepted: 23 May 2008 /
Published online: 13 June 2008

Abstract Principal component analysis has been applied to thirteen dimensionless geomorphic parameters for sixteen watersheds of the Chambal catchment of Rajasthan, India, in order to group the parameters under different components based on significant correlations. Results of the principal component analysis clearly revealed that first two principal components are strongly correlated with some of the geomorphic parameters. However, the third principal component is not found to be strongly correlated with any of the parameters but is moderately correlated with stream length ratio and bifurcation ratio. Furthermore, on the basis of the results, it is evident that some parameters are highly correlated with components but the parameters of hypsometric integral and drainage factor could not be grouped with any of the component because of its poor correlation with them. The principal component loadings matrix obtained using correlation matrix of ten parameters reveals that first three components together account for 87.01% of the total explained variance. Therefore, principal component lading matrix is applied in order to get

P. K. Singh (✉)
Department of Soil and Water Engineering,
College of Technology and Engineering,
Udaipur 313 001, India
e-mail: pksingh35@yahoo.com

V. Kumar
College of Technology and Engineering,
Udaipur, India

R. C. Purohit · M. Kothari
Department of SWE, CTAE Udaipur, India

P. K. Dashora
Department of Statistics, RCA, Udaipur, India

better correlations and clearly grouped the parameters in physically significant components.

Keywords Principal component analysis · Geomorphic parameters · Watershed · Rajasthan

1 Introduction

Hydrologic modeling is basically a tool for prediction of hydrologic behavior of a basin. Multivariate analysis is a collection of procedure for analyzing association between two or more set of variables that have been made on each object in one or more samples of object. Different approaches have been made in the past to intercorrelate the variables involving geomorphologic, geological, hydro-meteorological characteristics of the watersheds. Many of the geomorphic parameters are known to be strongly correlated. There is considerable amount of redundancy in the array of geomorphic parameters currently in use. The collection and analysis of geomorphological data are often time-consuming, there is clearly the need to reduce the number of parameters to a few that adequately simulate the drainage basin morphology. Such a screening will no doubt result in considerable saving in time and expenses involved in deriving geomorphological parameters from topographical maps. It will also provide a more rational basis for a multi-dimensional classification of drainage basins, which can form the basis for regional analysis and an objective selection of representative basins.

The screening of such a large number of interrelated variables for their underlying dimensions is best achieved by multivariate statistical techniques of principal component analysis. The principal component analysis breaks down the battery of intercorrelations among variables into a set of uncorrelated factors, these together summarize the data in the original matrix and explain the underlying relations and influences among the variables. The identification of these underlying dimensions will not only simplify future morphometric works but also provide criteria for an objective multi-dimensional morphological classification of drainage basins.

Principal component analysis is in fact an analysis of reduced space in which an attempt is made to find a smaller number of dimensions that retain most of the information in the original space. Therefore, in this study an attempt has been made to study the intercorrelationship (multicollinearity) among the variables in order to screen out the less significant variables out of the analysis and to arrange the remaining into physically significant groups by applying principal component analysis for better interpretability.

Principal component analysis by reducing a large number to a small number of principal components ensures economy in the use of large volume of data. A principal component conveys all essential information about the variables, ensuring economy in analysis and description while obtaining relatively accurate results (Sharma 2002).

It is evident from literature that multicollinearity exists between different geomorphic parameters. In the recent past few attempts have been made to solve the problem of interdependency among hydrologic and geomorphic parameters (Ravichandrana et al. 1996; Montes-Botella and Tenorio 2003; Mrklas et al. 2006; Magner and Brooks 2007).

In developing a model from mean annual flood in New England, Wong (1979) utilized a multivariate statistical technique and component analysis in analyzing the effects of twelve basins and climotological parameters. He found that principal components to main stream length and average slope are orthogonal to each other and at the same time covaried with mean annual flood, and therefore, selected new variables, i.e., principal components as surrogates of stream size and slope and used them as predictors of mean annual flood.

Bagley et al. (1964), Haan and Read (1970), Haan and Allen (1972), Decoursey and Deal (1974), and Pondzic and Trninic (1992) have also demonstrated the use of multiple regression analysis and principal component analysis for development of hydrological prediction equation involving geomorphic parameters. Mishra and Satyanatayana (1988) carried out principal component analysis with varimax rotation on the ten geomorphic parameters of Damodar Valley catchment of India and concluded that nine parameters could be significantly grouped into three components (steepness, shape and geological). Bifurcation ratio was found to be less significant in explaining the component variance. Kumar and Satyanarayana (1993) carried out principal component analysis for eastern red soil region of the India and concluded that circulatory ratio, ruggedness number and drainage factor have been found non significant for explaining the component variance. Tamene et al. (2006) applied principal component analysis to analyze the relationship between sediment yield and catchment characteristics and to determine the major factors controlling the variability of sediment yield for 11 catchments of northern Ethiopia. The results show that terrain form, gully erosion, surface lithology, and land cover explain most of the variability in sediment yield among the catchments.

1.1 Study Area and Data Source

Sixteen watersheds were chosen for the present analysis. These watersheds are located in Chambal catchment of Rajasthan, India (Fig. 1). The entire Chambal catchment lies between a longitude of 74°45′ to 75°50′ E and latitude of 23°30′ to 25°10′ N and spread over the states of Madhya Pradesh and Rajasthan. The Chambal river which originates from Vindhya hills flows into Madhya Pradesh and finally drains through Rajasthan. It has immense potentiality of water harvest. Four large dams, namely Gandhi Sagar in Madhya Pradesh, Ranapratap Sagar, Jawahar Sagar and Kota Barrage in Rajasthan, constructed on river Chambal, have a total catchment area of about 27000 sq. km. A large portion of catchment area, about 22850 sq. km is however in Madhya Pradesh and the rest of which is only about 4150 sq. km falls in the state of Rajasthan.

The climate of the region in general is of sub-humid with an annual average rainfall varying from about 650 mm to 850 mm. In general the topography of the region is undulating. The predominant slope in the hilly region ranges between 15 to 40 percent and in the valley portion it varies from almost level to 10 percent. Erosion problem in the catchments is quite intensive because of undulating topography,

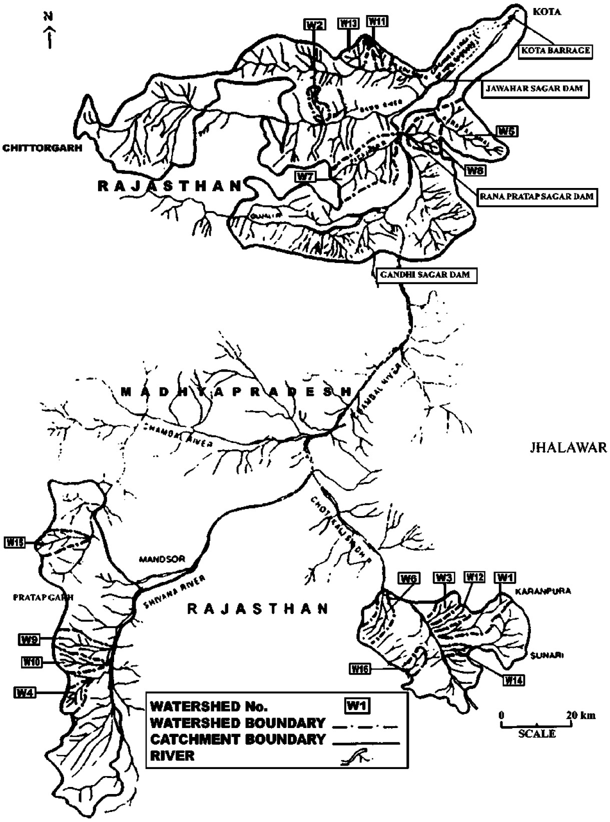

Fig. 1 Selected watersheds of the Chambal catchment

misuse of land and degradation of forest. Sheet and gully erosion are common in the area especially near the river.

2 Methodology

2.1 Geomorphic Parameters

Watershed characteristics play a vital role on the hydrologic responses of watersheds, and therefore, a number of parameters which signify the watershed characteristics are evaluated from the toposheets. The length parameters are measured with the help of a map measurer and area with the help of a planimeter. Ultimately these are used to obtain following 13 dimensionless parameters known as geomorphic parameters for the 16 watersheds of the Chambal catchment of Rajasthan, India. Singh (1992) and Singh (2000) also specify the important geomorphological characteristics of the watersheds. Thirteen salient parameters are selected in this study, which is based on the work conducted at Domodar Valley catchment, India (Kumar 1991).

1. Average slope of the watershed (S_a) is determined using the following relationship (Kumar 1991).

$$S_a = H\,L_{ca} \big/ (10\,A) \tag{1}$$

Where, S_a = average slope of the watershed, (%), H = maximum watershed relief (Elevation difference between most remote point to the outlet, m)
L_{ca} = Average length of the contour in km

$$= \sum_{i=1}^{n} L_{ci} \big/ n \tag{2}$$

Where, L_{ca} = length of each contour, km, n = number of identifiable contours, and A = drainage area of the watershed, sq km.

2. Relief ratio (R_r) is determined as the maximum watershed relief divided by the length of main stream.

$$R_r = \frac{H}{L} \tag{3}$$

Where, H = maximum watershed relief (Elevation difference between most remote point to the outlet, m) and L = length of main stream.

3. Relative relief (R_R) is determined as the ratio of the maximum watershed relief to the perimeter of the watershed.

$$R_R = \frac{H}{P} \times 100 \tag{4}$$

Where, H = maximum watershed relief (Elevation difference between most remote point to the outlet, m) and P = perimeter of the watershed.

4. Main stream channel slope (S_c), expressed in percent, is obtained by measuring the catchment area under the actual longitudinal profile of the main stream channel from gauge to divide using the following expression (Kumar 1991).

$$S_c = \frac{Area\ Under\ the\ curve}{5L^2} \tag{5}$$

Where, L= length of main stream, km.

5. Elongation ratio (R_e) is determined as the ratio between the diameter of a circle with the same area as the watershed and the maximum length of the watershed.

$$R_e = \frac{D_c}{L} \tag{6}$$

Where, D_c = diameter of a circle having same area as the watershed, L = maximum length of watershed.

6. Basin shape factor (S_b) is determined as the ratio between the square of the maximum length of the watershed and the area of the watershed.

$$S_b = \frac{\text{square of the maximum length of the watershed}}{\text{area of the watershed}} \tag{7}$$

7. Length–width ratio (L_b/L_w) is the ratio of the maximum length to the width of the watershed. Width is measured at the mid point of the longest stream length.

8. Stream length ratio (R_l) is the ratio between the mean length of a stream of a particular order and the mean stream length of the next higher order (Singh 2000).

$$R_l = \frac{\bar{L}_u}{\bar{L}_u - 1} \tag{8}$$

Where, $\bar{L}_u$ = mean length of stream of order u and $\bar{L}_u - 1$ = mean length of stream of next lower order.

9. Bifurcation ratio (R_b) is the ratio of the number of streams of a particular order to the number of streams of the next higher order (Singh 2000).

$$R_b = \frac{N_u}{N_u + 1} \tag{9}$$

Where, N_u = number of stream of u order, $N_u + 1$ = number of stream of $u + 1$ order.

10. Hypsometric analysis of drainage basin is carried out to develop the relationship between horizontal cross-sectional drainage basin area and the elevation. In analysis, a curve is derived by plotting the relative heights (h/H) and relative areas (a/A); the obtained curve is called as hypsometric curve (Suresh 1997). The shape of the hypsometric curve varies in early geologic stages of development of the drainage basin, but once a steady state is attained it tends to vary little despite lowering relief (Kumar 1991; Suresh 1997).

11. Circulatory ratio (R_c) is the ratio of circumference of a circle of same area as the watershed to the perimeter of the watershed.

$$R_c = \frac{A_u}{A_c} \tag{10}$$

Where, A_u = area of the watershed and A_c = area of circle having equal perimeter as the perimeter of the watershed.

12. Ruggedness number (R_N) is the drainage density times the maximum basin relief. The drainage density (D_d) is an important indicator of the linear scale of land-form elements in stream eroded topography and simply the ratio of total channel segment lengths cumulated for all orders within a basin to the basin area.

$$R_N = HD_d/1000 \tag{11}$$

Where, H = watershed relief and D_d = drainage density.

13. Drainage factor (D_f) is the ratio of stream frequency to the square of drainage density. Stream frequency (F_s) is the ratio of total number of streams of all order to the basin area.

$$D_f = \frac{F_s}{D_d^2} \tag{12}$$

The evaluated values of these geomorphic parameters are shown in Table 1.

Table 1 Selected dimensionless geomorphic parameters

Watershed	S_a	R_e	R_c	S_b	R_r	R_R	R_N	S_c	H_{si}	D_f	R_l	R_b	L_b/L_w
W1	1.850	0.532	0.731	4.493	0.0039	0.0017	0.082	0.204	0.541	0.608	2.062	3.580	2.848
W2	12.940	0.871	0.890	2.132	0.0481	0.0176	0.715	1.296	0.330	0.502	1.108	3.260	1.650
W3	1.625	0.782	0.756	2.079	0.0047	0.0014	0.064	0.314	0.580	0.718	1.177	3.651	1.398
W4	1.182	0.772	0.706	2.130	0.0033	0.0009	0.039	0.259	0.690	0.720	0.347	6.237	1.845
W5	6.178	0.422	0.591	7.125	0.0095	0.0042	0.486	0.397	0.625	0.653	1.917	3.664	6.295
W6	1.672	0.675	0.853	2.789	0.0041	0.0016	0.042	0.310	0.654	0.813	2.673	4.786	2.515
W7	2.744	0.791	0.742	2.002	0.0144	0.0043	0.319	0.944	0.494	0.677	2.452	5.042	1.655
W8	6.832	0.796	0.870	2.004	0.0019	0.0066	0.449	0.937	0.422	0.449	1.607	4.546	1.465
W9	0.622	0.776	0.892	2.110	0.0051	0.0019	0.049	0.375	0.540	1.175	0.873	3.359	1.694
W10	1.058	0.791	0.834	2.030	0.0052	0.0018	0.074	0.561	0.474	0.732	1.817	3.681	1.460
W11	1.302	0.784	0.892	2.071	0.0076	0.0027	0.046	0.946	0.638	1.202	0.958	4.123	1.567
W12	2.771	0.760	0.923	2.200	0.0082	0.0032	0.051	0.478	0.481	0.570	1.463	2.645	1.641
W13	2.830	0.746	0.876	2.286	0.0094	0.0035	0.115	0.601	0.448	0.748	0.663	3.207	1.793
W14	0.593	0.633	0.753	3.174	0.0056	0.0021	0.038	0.336	0.237	0.699	1.068	7.998	2.181
W15	1.126	0.994	0.728	1.287	0.0058	0.0014	0.042	0.560	0.472	0.748	1.041	2.801	1.408
W16	2.088	0.949	0.904	1.295	0.0131	0.0040	0.054	0.592	0.369	0.755	1.207	2.654	0.933

2.2 Principal Component Analysis

The method of principal components or component analysis is based upon the early work of Pearson with the specific adaptations to principal component analysis suggested by the work of Hotelling (1933). The geomorphic parameters are usually many times correlated. The correlation indicates that some of the information contained in one variable is also contained in some of the other remaining variables. More specifically, the first principal component is that linear combination of the original variables which contributes a maximum to their total variance; the second principal component, uncorrelated with the first, contributes a maximum to the residual variance, and so on until the total variance is analyzed. Since the method is so dependent on the total variance of the original variables, it is most suitable when all the variables are measured in the same units. Hence, it is customary to express the variables in standard form, i.e., to select the unit of measurement for each variable so that its sample variance is one. Then, the analysis is made on the correlation matrix, with the total variance equal to n. The objectives are achieved in two steps:

Step 1 Calculate the correlation matrix, *R*
Step 2 Calculate the principal component loading matrix by principal component analysis.
Step 3 In the principal component (PC) loading matrix, Eigen value greater than one indicates significant PC loading.

Eigen value indicated how well each of the identified factors fit the data from all the geomorphic parameters on all the principal components.

2.3 Correlation Matrix

The inter-correlation matrix of the geomorphic parameters is obtained by using the following procedure:

1. The parameters are standardized:

$$X = \left(x_{ij} - x_j\right)/S_j \tag{13}$$

   where

   x denotes the matrix of standardized parameters
   x_{ij} ith observation on jth parameter
   i 1, ..., N (Number of Observations)
   j 1, ..., P (Number of Parameters)
   x_j mean of the jth parameter
   S_j standard deviation of the jth parameter

2. The correlation matrix of parameters is the minor product moment of the standardized predictor measures divided by N and is given by

$$\mathrm{R} = \left(\mathrm{x}' \times \mathrm{x}\right)/\mathrm{N} \tag{14}$$

   where, x' denotes the transpose of the standardized matrix of predictor parameters.

2.4 Principal Component Loading Matrix

The principal component loading matrix which reflects how much a particular parameter is correlated with different factors, is obtained by premultiplying the characteristic vector with the square root of the characteristic values of the correlation matrix.

$$\text{Thus, } A = Q \times D^{0.5} \tag{15}$$

where

- A principal component loading matrix,
- Q characteristic vector of the correlation matrix,
- D characteristic value of the correlation matrix

3 Results and Discussion

The correlation matrix (Table 2) of the thirteen selected geomorphic parameters reveals that strong correlations (correlation coefficient more than 0.9) exist between basin shape factor (S_b) and length width ratio (L_b/L_w), between average slope of the watershed (S_a) and ruggedness number (R_N), between average slope of the watershed (S_a) and relative relief (R_R), between relief ratio (R_r) and relative relief and elongation ratio (R_e) and basin shape factor (S_b). Also, good correlations (correlation coefficient more than 0.75) exist between R_R and R_N, S_a and R_r and between R_r and main stream channel slope (S_c). Some more moderately correlated parameters (correlation coefficient more than 0.6) are S_a with S_c, R_r with R_N, R_r with S_c and R_N with S_c. It is very difficult at this stage to group the parameters into components and attach any physical significance because some parameters like D_f and H_{si} do not show any significant correlation with any of the parameters. Hence, in the next step, the principal component analysis has been applied. The correlation matrix is subjected to the principal component analysis.

The principal component loading matrix obtained from correlation matrix (Table 3) reveals that the first three components whose Eigen values are greater than one, together account for about 77.26% of the total explained variance. The first component is strongly correlated (loadings of more than 0.8) with relative relief, relief ratio, average slope and main stream channel slope and moderately (loadings of more than 0.6) with ruggedness number, which may be termed as a slope or steepness component. The second component is strongly correlated with basin shape factor, length width ratio and elongation ratio but moderately with circulatory ratio and can be termed as shape component. The third component does not strongly correlate with any geomorphic parameters but moderately correlates with stream length ratio and bifurcation ratio (loadings of more than 0.6) and may be termed as drainage component. It is evident from these results that some parameters are highly correlated with components but the parameters H_{si} and D_f could not be grouped with any of the components because of its poor correlation with them.

In order to screen out parameters having less significance in explaining the component variance, the parameter D_f is first screened out from the analysis. Then, the correlation matrix and principal component loading matrix are obtained for

Table 2 Intercorrelation matrix of the selected geomorphic parameters

Parameter	S_a	R_e	R_c	S_b	R_r	R_R	R_N	S_c	H_{si}	D_f	R_l	R_b	L_b/L_w
S_a	1.000	0.005	0.094	0.179	0.801	**0.944**	**0.945**	0.676	−0.294	−0.547	0.124	−0.212	0.211
R_e	0.005	1.000	0.519	−**0.915**	0.251	0.196	−0.107	0.453	−0.293	0.096	−0.263	−0.304	−0.812
R_c	0.094	0.519	1.000	−0.643	0.226	0.278	−0.103	0.406	−0.299	0.214	−0.096	−0.338	−0.657
S_b	0.179	−**0.912**	−0.643	1.000	−0.080	−0.041	0.299	−0.340	0.254	−0.169	0.276	0.107	**0.952**
R_r	0.801	0.251	0.226	−0.080	1.000	**0.923**	0.705	0.691	−0.408	−0.275	−0.031	−0.223	−0.051
R_R	**0.944**	0.196	0.278	−0.041	**0.923**	1.000	0.859	0.778	−0.450	−0.422	0.010	−0.189	−0.023
R_N	**0.945**	−0.107	−0.103	0.299	0.705	0.859	1.000	0.669	−0.240	−0.519	0.226	−0.098	0.345
S_c	0.676	0.453	0.406	−0.340	0.691	0.778	0.669	1.000	−0.359	−0.141	0.056	−0.185	−0.276
H_{si}	−0.294	−0.293	−0.299	0.254	−0.408	−0.450	−0.240	−0.359	1.000	0.388	0.315	−0.090	0.323
D_f	−0.547	0.096	0.214	−0.169	−0.275	−0.422	−0.519	−0.141	0.388	1.000	−0.221	−0.029	−0.114
R_l	0.124	−0.263	−0.096	0.276	−0.031	0.010	0.226	0.056	0.315	−0.221	1.000	−0.270	0.271
R_b	−0.212	−0.304	−0.338	0.107	−0.223	−0.189	−0.098	−0.185	−0.090	−0.029	−0.270	1.000	0.080
L_b/L_w	0.211	−0.812	−0.657	**0.952**	−0.051	−0.023	0.345	−0.76	0.323	−0.114	0.271	0.080	1.000

Figures in bold face indicate strong correlations

Table 3 Principal component loading matrix of selected geomorphic parameters

Parameters	Principal components												
	1	2	3	4	5	6	7	8	9	10	11	12	13
S_a	0.882	0.418	0.032	0.053	0.016	0.043	0.173	−0.062	0.059	−0.034	0.045	−0.027	0.007
R_e	0.392	−0.817	0.077	−0.030	0.082	−0.345	−0.113	−0.015	0.151	0.048	0.034	0.020	0.001
R_c	0.384	−0.675	0.246	0.064	−0.164	0.490	0.238	−0.056	0.059	0.057	−0.013	0.013	0.001
S_b	−0.234	0.934	0.002	0.124	−0.194	0.105	−0.051	−0.022	−0.048	0.005	0.056	0.033	0.003
R_r	0.883	0.107	−0.007	0.225	−0.056	−0.063	−0.047	0.377	−0.047	0.059	−0.030	0.004	0.003
R_R	0.965	0.181	−0.033	0.128	−0.007	0.032	0.083	0.086	0.018	−0.041	0.039	0.001	0.003
R_N	0.794	0.551	0.013	0.052	0.139	−0.052	−0.010	−0.145	0.063	−0.110	−0.063	0.019	−0.011
S_c	0.857	−0.137	0.111	0.159	0.307	0.060	−0.176	−0.226	−0.142	0.092	0.007	−0.002	0.001
H_{si}	−0.539	0.218	0.566	0.240	0.331	−0.252	0.331	0.024	−0.034	0.030	−0.002	−0.007	−0.000
D_f	−0.443	−0.397	0.268	0.707	0.065	0.139	−0.203	0.032	0.051	−0.084	0.011	−0.006	0.000
R_l	0.016	0.391	0.653	−0.468	0.299	0.253	−0.166	0.135	0.047	−0.016	0.012	−0.003	0.000
R_b	−0.272	0.131	−0.756	0.073	0.523	0.209	0.069	0.075	0.062	0.024	0.010	0.004	0.001
L_b/L_w	−0.204	0.920	0.060	0.201	−0.122	−0.010	−0.089	−0.072	0.145	0.136	−0.023	−0.011	−0.002
Eigen value	4.813	3.757	1.474	0.945	0.679	0.574	0.333	0.259	0.089	0.060	0.014	0.003	0.000

Table 4 Principal component loading matrix of ten finally screened out geomorphic parameters

Parameters	Principal components									
	1	2	3	4	5	6	7	8	9	10
S_a	0.938	0.255	0.054	−0.048	0.019	−0.202	−0.024	0.075	0.008	0.037
R_e	0.258	−0.917	−0.115	−0.038	0.084	−0.029	0.260	−0.000	0.051	−0.000
S_b	0.069	0.980	0.021	−0.141	0.018	0.063	−0.025	−0.038	0.088	0.006
R_r	0.904	−0.044	0.106	−0.136	−0.332	0.187	0.072	−0.029	−0.016	0.018
R_R	0.979	0.023	0.125	−0.033	−0.116	−0.054	−0.057	0.038	0.023	−0.051
R_N	0.883	0.386	0.063	0.102	0.152	−0.133	0.058	−0.110	−0.029	−0.004
S_c	0.849	−0.276	0.005	0.236	0.281	0.240	−0.099	0.019	0.006	0.005
R_l	0.099	0.386	−0.734	0.531	−0.134	−0.006	0.044	0.010	0.006	−0.000
R_b	−0.285	0.148	0.792	0.509	−0.065	−0.006	0.075	0.012	0.012	0.002
L_b/L_w	−0.021	0.950	0.005	−0.151	0.129	0.098	0.205	0.060	−0.035	−0.011
Eigen value	4.319	3.168	1.214	0.674	0.272	0.169	0.140	0.026	0.014	0.004

12 parameters. The less correlated parameters such as R_c from second component and H_{si} from third component are also screened out because of lower principal component loading matrix and same analysis is repeated with only 10 variables.

The principal component loadings matrix obtained using the correlation matrix of 10 parameters (Table 4) reveals that the first three components now together accounts for 87.01% of the total explained variance showing an increase of about 9.71%. The principal component loadings here also improved considerably in almost all significant parameters. The relative relief, relief ratio and average slope are highly correlated (loadings of more than 0.9) with the first component. The ruggedness number and main stream channel slope are also having good correlations (loadings of more than 0.8) with the first component. The basin shape factor, length width ratio and elongation ratio are highly correlated (loadings of more than 0.9) with second component. The correlations of R_l and R_b with third component are observed to increase significantly.

The results of this study reveal that steepness component is the dominant component followed by shape and drainage components. Therefore, it can be presumed that hydrologic response like runoff yield and soil loss for these watersheds will be high. The peak runoff rate of the watershed will not be achieved frequently only because of less dominancy of shape and drainage components. Some watersheds (i.e., W4, W7 and W14) exhibited higher bifurcation ratios, (R_b greater than 4), which would result lower but extended peak flow, whereas remaining watersheds with lower R_b will produce sharp peak flow. Furthermore, most of the watersheds (except watersheds W4, W9, and W13) exhibited higher values of elongation ratio (greater than 0.8), which shows that the area is having steep ground slope.

It is observed that the first component is strongly correlated with relative relief, average slope of the watershed, ruggedness number, relief ratio and main stream channel slope which are grouped under slope or steepness component. The second component is strongly correlated with basin shape factor, length width ratio and elongation ratio of the watershed and is termed as shape component. The third component is strongly correlated with bifurcation ratio and stream length ratio and hence is called as drainage component.

It can be seen how useful the principal component analysis have been in screening out the parameters or variables of least significance and in regrouping the remaining variables into physically significant factors. Multiple regression techniques can then applied in modeling the hydrologic responses such as runoff and sediment yields from the watersheds. One parameter each from significant components may form a set of independent parameters at a time in modeling the said hydrologic responses.

4 Conclusions

In the present study, thirteen geomorphic parameters of sixteen watersheds located in Chambal catchment of Rajasthan, India were chosen for the analysis. The correlation matrix of the thirteen selected geomorphic parameters revealed that strong correlation (correlation coefficient>0.9) exist between basin shape factor and length width ratio, between average slope of the watershed and the ruggedness number, between average slope of the watershed and relative relief, between relief ratio and relative relief, and elongation ratio and basin shape factor. The principal component loading matrix obtained from correlation matrix reveals that the first three components, whose Eigen values are greater than one, together accounts for about 77.26% of the total explained variance. Based on the results of the principal component analysis, first component is strongly correlated with relative relief, relief ratio, average slope of the watershed and main stream channel slope. The second component is strongly correlated with basin shape factor, length width ratio, and elongation ratio. However, third component is not found strongly correlated with any of the geomorphologic parameters but moderately correlated with stream length ratio, and bifurcation ratio. The hypsometric integral and drainage factor could not be grouped with any of the components because of its poor correlations with them. After screening out the hypsometric integral, drainage factor and circulatory ratio, the principal component loadings matrix of ten parameters indicate that first three components together account for 87.01% of the total explained variance. Based on the properties of the geomorphic parameters, three principal components were defined as steepness, shape, and drainage components. Moreover, it is concluded that in modeling the hydrologic responses such as runoff and sediment yield from small watersheds, the principal component analysis is good tool for screening out the insignificant parameters from the analysis.

References

Bagley JM, Jeppson RW, Milligan CH (1964) Water yields in Utah agricultural experimental station. Utah State University, Logan, Utah Special Report No. 18

Decoursey DG, Deal RB (1974) General aspect of multivariate analysis with applications to some problems in hydrology. In: Proceedings of symposium on statistical hydrology, USDA, miscellaneous publication No. 1275. Washington DC, pp 47–68

Haan CT, Allen DM (1972) Comparison of multiple regression and principal component regression for predicting water yields in Kentucky. Water Resour Res 8(6):1593–1596

Haan CT, Read HR (1970) Prediction of monthly seasonal and annual runoff volumes for small agricultural watersheds in Kentucky, Bulletin 711, Kentucky Agricultural Experiment Station, University of Lexington, Kentucky

Hotelling H (1933) Analysis of a complex of statistical variables into principal components. J Educ Psychol 24:417–441, 498–520

Kumar V (1991) Hydrologic response models for prediction of runoff and sediment yields from small watersheds. Unpublished Ph.D. Thesis. Indian Institute of Technology, Kharagpur, India, p 350

Kumar V, Satyanarayana T (1993) Application of principal component analysis in grouping geomorphic parameters for hydrologic modeling. J Indian Water Resour Soc 13(1–2):50–59

Magner JA, Brooks KN (2007) Predicting stream channel erosion in the lacustrine core of the upper Nemadji River, Minnesota (USA) using stream geomorphology metrics. Environ Geol, published online, http://www.springerlink.com/content/e65ju8857175kh23/?p=2dcbc28f160c4bfc86a413d59b4b1e82&pi=0. Accessed 8 March 2008

Mishra N, Satyanatayana T (1988) Parameter grouping—a prelude to hydrologic modeling. Indian J Power River Val Dev 256–260, September

Montes-Botella C, Tenorio MD (2003) Water characterization and seasonal heavy metal distribution in the Odiel River (Huelva, Spain) by means of principal component analysis. Arch Environ Contam Toxicol 45(4):436–444

Mrklas O, Bentley LR, Lunn SRD, Chu A (2006) Principal component analyses of groundwater chemistry data during enhanced bioremediation. Water Air Soil Pollut 169(1–4):395–411

Pondzic K, Trninic D (1992) Principal component analysis of a river basin discharge and precipitation anomaly fields associated with the global circulation. J Hydrol 132(1–4):343–360

Ravichandrana S, Ramanibai R, Pundarikanthan NV (1996) Ecoregions for describing water quality patterns in Tamiraparani basin, South India. J Hydrol 178(1–4):257–276

Sharma KR (2002) Research methodology. National Publishing House, New Delhi, p 514

Singh VP (1992) Elementary hydrology. Prentice-Hall of India Private Limited, New Delhi, India, p 973
Singh RV (2000) Watershed planning and management. Yash Publishing House, Bikaner, p 470
Suresh R (1997) Soil and watert conservation engineering. Standard Publishers Distributors, New Delhi, p 973
Tamene L, Park SJ, Dikau R, Vlek PLG (2006) Analysis of factors determining sediment yield variability in the highlands of northern Ethiopia. Geomorphology 76(1–2):76–91
Wong ST (1979) A multivariate statistical model for predicting mean annual flood in New England. Annals Assoc Am Geographers 53:293–311