RESEARCH PAPER International Journal of Agricultural Engineering | Volume 13 | Issue 1 | April, 2020 | 62-66
ISSN–0974–2662 Visit us : *www.researchjournal.co.in* **DOI: 10.15740/HAS/IJAE/13.1/62-66**

Geomorphic modelling for small watersheds using Principal Component Analysis

Sandip Nikam

Received : 10.01.2020; **Revised :** 15.02.2020; **Accepted :** 01.03.2020

Author for Correspondence:
Sandip Nikam
Department of Agricultural Engineerng, College of Agriculture (MPKV), **Dhule (M.S.) India**
Email: spnikam74@gmail.com

ABSTRACT : Principal Component Analysis was carried out for grouping the different parameters into the Principal Components. To understand the behaviour of all the parameters pertaining to study areas, and to reduce the dimensionality of database, the data pertaining to twelve parameters of ten small watersheds were submitted for Principal Component Analysis. The method of components analysis, then, involves the rotation in the total variable space - an orthogonal or uncorrelated transformation wherein each of the n original variables is describable, in terms of the n new principal components. An important feature of the new components is that they account, in turn, for a maximum amount of variance of the variables. Analysis extracted three components as a Principal Components with 10 parameters, accounting for a total variance of 97.256 per cent. The first component is highly correlated with R_e, R_c, S_b and L_{bw} accounting for 68.52 per cent variance. Second component is strongly correlated with R_N accounting for 18.60 per cent variance and Third with S_c, accounting for 10.13 per cent variance. Finally, these extracted 10 parameters were used for modeling for prediction of sediment yield and runoff from selected small watersheds of Tapi basin, Maharashtra, India.

KEY WORDS : Principal Component analysis, Geomorphological parameters, Morphometric model, Small watersheds

HOW TO CITE THIS PAPER : Nikam, Sandip (2020). Geomorphic modelling for small watersheds using Principal Component Analysis. *Internat. J. Agric. Engg.*, **13**(1) : 62-66, **DOI: 10.15740/ HAS/IJAE/13.1/62-66.**

One of the most important tools in hydrological analysis is the morphometric survey of the watersheds, which allows establishing evaluation parameters on the behaviour of the hydrological system of the basin area. This study, when properly combined with analysis of geomorphological parameters by Principal Components Analysis, helps to establish hydrological models for prediction of sediment production rate and runoff from the basin area. Therefore, in this study an attempt has been made to study the intercorrelationship 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 along with the orthogonal rotation for better interpretability.

Haan and Read (1970); Haan and Allen (1972); Decoursey and Deal (1974) and Pondzic and Trninic (1992) have demonstrated the use of multiple regression analysis and principal component analysis for development of hydrological prediction equation involving geomorphic parameters. 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.

Singh *et al.* (2009) used Principal Component Analysis to screen out the less correlated parameters and to regroup the correlated parameters into physically significant components. They found the out of thirteen geomorphological parameters, three parameters were not correlated with others and therefore screened out to regroup remaining ten parameters into three principal components.

The study area is Tapi Basin which is situated between 68°302 to 70°452 E longitudes and 22°180 to 23°252 N latitude. The Tapi River basin covers an area of 65,145 km² that makes up almost two per cent of the total area of India. The basin mostly lie in the northern and eastern districts Maharashtra state, including places like Dhule, Jalgaon, Nashik, Nandurbar, Amravati, Akola, Washim, and Buldhana districts. The river receives discharge from 14 main tributaries, 4 on the right bank and 10 on the left bank, of which the Purna River, Girna River, Panzara River, Waghur River, Bori River, Amarwati river, Mousam river and Aner River are the most important.

Various watersheds in the area of interest were marked using the Survey of India (SOI) toposheets. For the preparation of the drainage and contour maps at higher scale, digitized toposheets at the scale 1: 2,50,000 and undigitized toposheets at the scale 1: 50,000 were used which were digitized later. ArcGIS 9.3 software was used to evaluate the twelve geomorphological parameters of the selected ten watersheds from digitized toposheets.

■ METHODOLOGY

Geomorphological 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. Singh (1992) and Singh (2000) also specified the important geomorphological characteristics of the watershed. Twelve salient parameters are selected in this study for Tapi basin of Maharashtra state, India.

Table A :

Sr. No.	Geomorphological parameter	Formula
S_a	Avg. slope of the watershed	$S_a = \frac{H\sum_{i=1}^{n} L_{ci}}{10nA}$
R_e	Elongation ratio	$R_e = \frac{2\sqrt{A}}{L_b\sqrt{\text{®}}} = 1.12838\sqrt{A/L_b}$
R_c	Circulatory ratio	$R_c = \frac{2\sqrt{\text{®} A}}{L_p} = 3.544\frac{\sqrt{A}}{L_p}$
S_b	Basin shape Factor	$S_b = L_b^{\ 2}/A$
R_f	Relief ratio	$Rf = \frac{H}{L_b}$
R_r	Relative ratio	$R_r = \frac{H}{L_p}$
R_n	Ruggedness number	$R_n = \frac{H\ \ D_d}{1000}$
S_c	Main stream channel slope	$S_c = \frac{Area\ Under\ the\ curve}{5L^2 ms}$
D_f	Drainage factor	$D_f = F_s / D_d^{\ 2}$
R_l	Stream length ratio	$\log_{10}\bar{L}_u = a + b\,u\ ;\ R_1 = anti\log b$
R_b	Bifurcation ration	$\log_{10} N_u = a - b\,u\ ;\ R_b = Anti\log b$
L_{bw}	Length width ratio	L_b/L_w

Principal component analysis:

The principal Components Analysis with rotations was carried out in following three steps:

Step 1 - Calculate the correlation matrix, R

Step 2 - Calculate the unrotated factor loading matrix by principal component analysis.

Step 3 - Calculate the rotated factor loading matrix to enhance interpretability by orthogonal transformation.

SPSS 16.0 software have been used for obtaining correlation matrix, first (unrotated) factor loading matrix, orthogonal rotation of a factor loading matrix using a generalized orthomax criteria including quartimax, varimax, and equamax. The varimax method attempts to load highly a relatively low number of variables on each factor.

Correlation matrix:

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

(i) The parameters are standardized:

$$\mathbf{X} = \frac{(\mathbf{x_{ij}} - \mathbf{x_j})}{\mathbf{S_j}}$$

where, x denotes the matrix of standardized parameters, x_{ij} = i^{th} observation on j^{th} parameter

i = 1, ……, N (no. of observations)

j = 1, ……, P (no. of parameters)

x_j = Mean of the jth parameter

S_j = Standard deviation of the j^{th} parameter

(ii) The correlation matrix of predictor parameters is the minor product moment of the standardized predictor measures divided by N and is given by

$$\mathbf{R} = \frac{(\mathbf{x' * x})}{\mathbf{N}}$$

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

First factor loading matrix:

The unrotated or first factor 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.

Thus, $\mathbf{A = Q * D^{0.5}}$

where, A = First factor loading matrix,

Q = Characteristic vector of the correlation matrix

D = Characteristic value of the correlation matrix

Rotated factor loading matrices:

When a transformation matrix is post-multiplied to the first factor loading matrix, the rotated loading matrix is obtained. Hence,

$\mathbf{B = A * H}$

where, B = Rotated factor loading matrix,

H = Transformation matrix

While deriving the rotated factor loading matrix only those components whose eigen-values are greater than one are retained.

RESULTS AND DISCUSSION

The inter correlation matrix (Table 1) was developed using twelve selected geomorphic parameters of the ten watersheds. It reveals that strong correlations (correlation co-efficient more than 0.9) exist between Re and Sb, between Re and Lbw, between Sb and Lbw and between Rf and Rr. Also, good correlations (correlation co-efficient more than 0.75) exist between Re and Rc, Rc and Sb, Rc anf Lbw and between Df and Rl. Some more moderately correlated parameters (correlation co-efficient more than 0.6) are Sb with Rf, Rf with Rl, Rf with Lbw, Rr with RN, Rr with Rl and Sc with Df. It is very difficult at this stage to group the parameters into components and attach any physical significance because some parameters like Sa and Rb 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 of 12 parameters (Table 2) reveals that the first three components whose Eigen values are greater than one, together account for about 92.36 per cent of the total explained variance. The first component is strongly correlated (loadings of more than 0.9) with R_e, S_b, R_l and L_{bw} but moderately (loadings of more than 0.7) with Sa. The second component is strongly correlated with R_N. The third component does not strongly correlate with any geomorphic parameters but moderately correlates with S_c.

It is observed from Table 2 that some parameters have high, good or moderate correlation with components but the parameter R_b could not be grouped with any one of the components because of its poor correlation (0.4 to 0.5) with them. Therefore in the second step, the parameter R_b was first screened out and remaining 11

Table 1 : Intercorrelation matrix of the selected geomorphic parameters

Parameters	Sa	Re	Rc	Sb	Rr	Rf	R_N	Sc	Df	Rl	Rb	Lbw
Sa	1.000	0.110	0.205	-0.255	0.354	0.462	0.315	0.331	0.323	-0.409	0.248	-0.284
Re	0.110	1.000	0.825	-0.976	0.575	0.441	-0.047	0.008	0.496	-0.576	0.339	-0.951
Rc	0.205	0.825	1.000	-0.855	0.226	0.205	-0.275	-0.031	0.488	-0.434	0.375	-0.840
Sb	-0.255	-0.976	-0.855	1.000	-0.613	-0.514	-0.053	-0.034	-0.416	0.539	-0.392	0.974
Rr	0.354	0.575	0.226	-0.613	1.000	0.964	0.598	0.128	0.147	-0.637	0.098	-0.607
Rf	0.462	0.441	0.205	-0.514	0.964	1.000	0.658	0.134	0.111	-0.635	0.061	-0.511
R_N	0.315	-0.047	-0.275	-0.053	0.598	0.658	1.000	-0.025	-0.361	-0.072	-0.086	0.023
Sc	0.331	0.008	-0.031	-0.034	0.128	0.134	-0.025	1.000	0.679	-0.172	0.086	-0.080
Df	0.323	0.496	0.488	-0.416	0.147	0.111	-0.361	0.679	1.000	-0.813	0.088	-0.469
Rl	-0.409	-0.576	-0.434	0.539	-0.637	-0.635	-0.072	-0.172	-0.813	1.000	0.078	0.589
Rb	0.248	0.339	0.375	-.392	0.098	0.061	-0.086	0.086	0.088	0.078	1.000	-0.358
Lbw	-0.284	-0.951	-0.840	0.974	-0.607	-0.511	0.023	-0.080	-0.469	0.589	-0.358	1.000

Table 2 : Principal component loading matrix of selected geomorphic parameters

Parameters	Principal components											
	1	2	3	4	5	6	7	8	9	10	11	12
Sa	0.767	0.204	0.394	-0.289	0.362	-0.019	-0.010	-0.003	0.001	0.000	0.000	0.000
Re	0.955	-0.100	-0.233	0.149	-0.037	0.019	-0.022	-0.006	0.003	0.000	0.000	0.000
Rc	0.897	-0.296	-0.281	0.122	0.108	0.005	0.049	-0.010	0.001	0.000	0.000	0.000
Sb	-0.965	0.055	0.185	-0.177	-0.020	-0.011	0.011	0.014	-0.003	0.000	0.000	0.000
Rf	0.866	0.473	0.053	0.080	-0.125	-0.040	-0.018	0.008	0.001	0.000	0.000	0.000
Rr	0.827	0.543	0.103	0.032	-0.082	-0.052	0.020	0.013	0.000	0.000	0.000	0.000
R_N	0.152	0.926	0.287	0.171	0.003	0.087	0.009	-0.005	0.000	0.000	0.000	0.000
Sc	0.369	-0.447	0.790	-0.088	-0.18	-0.003	0.008	-0.028	0.001	0.000	0.000	0.000
Df	0.820	-0.341	-0.123	-0.434	-0.064	0.058	-0.007	0.015	0.000	0.000	0.000	0.000
Rl	-0.929	-0.114	0.091	0.320	0.117	0.000	-0.013	-0.013	0.001	0.000	0.000	0.000
Rb	0.575	-0.572	0.413	0.410	0.055	0.014	-0.003	0.037	-0.001	0.000	0.000	0.000
Lbw	-0.973	0.076	0.172	-0.131	-0.013	0.010	0.014	0.024	0.006	0.000	0.000	0.000
Eigen value	7.638	2.18	1.266	0.672	0.22	0.016	0.004	0.004	0.000	0.000	0.000	0.000

Table 3 : Principal component loading matrix of final geomorphic parameters

Parameters	Principal components									
	1	2	3	4	5	6	7	8	9	10
Re	0.974	-0.106	-0.165	0.103	0.003	-0.040	0.007	-0.001	0.000	0.000
Rc	0.906	-0.314	-0.241	0.137	0.039	0.057	0.003	0.000	0.000	0.000
Sb	-0.975	0.049	0.162	-0.140	-0.013	0.015	0.000	-0.001	0.000	0.000
Rf	0.881	0.463	0.071	-0.013	-0.06	-0.018	0.002	0.001	0.000	0.000
Rr	0.840	0.527	0.103	-0.046	-0.051	0.035	0.000	0.001	0.000	0.000
R_N	0.155	0.978	0.103	0.007	0.095	-0.004	0.000	0.000	0.000	0.000
Sc	0.286	-0.250	0.907	0.184	0.005	0.002	0.000	0.000	0.000	0.000
Df	0.818	-0.48	0.146	-0.276	0.051	-0.014	0.000	0.002	0.000	0.000
Rl	-0.955	0.024	-0.126	0.268	0.009	-0.012	0.000	0.002	0.000	0.000
Lbw	-0.981	0.081	0.128	-0.119	0.005	0.011	0.012	0.001	0.000	0.000
Eigen value	6.852	1.86	1.013	0.247	0.02	0.007	0.000	0.000	0.000	0.000

parameters are subjected to the principle component analysis. It reveals from the principle component loading matrix obtained from correlation matrix of 11 parameters that each parameter is having high, good or moderate correlation with first, second or third component. Further they are subjected different methods of transformation (rotation) of the first factor loading matrix such as varimax, equamax and quartimax. It is observed in the rotated component matrix by varimax method of the three principle components that the parameter Sa could not be grouped with any one of the components because of its poor correlation (0.4 to 0.5) with them. The parameter Sa is therefore screened out in the next step for PCA and the same analysis is repeated with only 10 variables.

The first factor loadings matrix obtained using the correlation matrix of 10 parameters (Table 3) reveals that the first three components now together accounts for 97.25 per cent of the total explained variance showing an increase of about 4.89 per cent. The first factor loadings here also improved considerably in almost all significant parameters. The R_e, R_c, S_b and L_{bw} are highly correlated (loadings of more than 0.9) with the first component. The R_N is highly correlated with second component. The third component is highly correlated with S_c.

The analytical rotations were carried out for the components having Eigen value more than one in order to redistribute the explained variance in improving the factor loadings. All the transformations almost resulted in the same loading trends.

It can be seen how useful the factor analysis and 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.

■ REFERENCES

Decoursey, D.G. and Deal, R.B. (1974). General aspect of multivariate analysis with applications to some problems in hydrology. Proceedings of symposium on statistical hydrology, USDA, miscellaneous publication No. 1275, Washington D.C.: 47-68.

Haan, C.T. and Allen, D.M. (1972). Comparison of multiple regression and principal component regression for predicting water yields in Kentucky. *Water Resources Research,* **8**(6): 1593-1596.

Haan, C.T. and Read, H.R. (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.

Pondzic, K. and 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):343–36.

Kumar, V. and 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

Singh, P.K., Virendra, K., Purohit, R.C., Kothari, M. and Dashora, P.K. (2009). Application of principle component analysis in grouping geomorphic parameters for hydrological modelling. *Water Res. Management,* **23**: 325-339

Singh, R.V. (2000). Watershed planning and management. Yash Publishing House, Bikaner: 470.

Singh, S. (1992). *Quantitative Geomorphology of the Drainage Basin, Readings on Remote Sensing Applications*, T. S. Chouhan and K. N. Joshi, Eds., Scientific Publishers, Jodhpur, India.