The Andhra Agric. J 61(2): 402-407, 2014

Principle Component Approach for Clustering the Rainfall Pattern in Visakhapatnam District

G Samba Siva, V Srinivasa Rao, D Ratna Babu and V Sekhar
Department of Statistics and Mathematics, Agricultural College, Bapatla 522 101

ABSTRACT

A multivariate approach based on principle component analysis has been proposed to study the rainfall based classification of different mandals of Visakhapatnam district of Andhra Pradesh. The approach was applied to study the rainfall patterns of 42 mandals based on 25 years of rainfall data of Visakhapatnam district from 1986-2010. The district exhibits variability in monthly and mandal wise rainfall and the first six principle components explains 74.22 per cent of rainfall variability which are statistically significant. The application of this approach identified medium rainfall (862 mm-1162 mm) was the most frequent representative pattern of rainfall in all mandals of Visakhapatnam district. The analysis based on multivariate approach like principle component analysis provides useful information about the rainfall patterns that are likely to occur in different regions in different periods.

Key words : Coefficient of variation, Common principle components, Principle component analysis, Rainfall.

Rainfall plays an important role in agricultural planning, particularly in rainfed agriculture. Rainfall has been studied at length in several studies (Hills and Morgan, 1981, Kulkarni and Pandit, 1988, Prabhakaran *et al.,* 1992, Kulkarni and Reddy, 1994, Kulkarni and Rao, 2000). These studies cover important topics such as probabilistic estimation of rainfall by fitting statistical distributions, rainfall based classification of regions and to study of crop yield, rainfall and weather relations.

Classification of regions which accounts for due consideration of temporal variation in the rainfall and the amount of rainfall would help in planning suitable crop strategies such as date of sowing, fertilizer application, plant protection measures and suitable crop variety can be planned which would be specific for these homogenous regions. In the context of climatic classification on the basis of rainfall clustering was carried out under single sample situation, based on mean values of rainfall variates. Such a summarization is not meaningful as rainfall exhibits a lot of year to year variation. Hence, clustering under multiple sample situations is to be preferred. When clustering methods are to be applied to this data situation, it is essentially to verify the assumption of homogeneity among the covariance matrices of the centers. There seems to no serious thought on this crucial assumption in most of the applications.

The district receives annual normal rainfall of 1202 mm, of which South-West monsoon accounts for 72 per cent of the normal while North-East monsoon contributes 13.9 per cent of the normal rainfall. The rest is shared by summer and winter rains.

MATERIAL AND METHODS

SOURCE AND NATURE OF DATA

The present study is based on 25 years of mandalwise rainfall data of Visakhapatnam district covering the years 1986-2010. The relevant data were collected from the office of Chief Planning Office (CPO), Visakhapatnam.

PRINCIPLE COMPONENT ANALYSIS

Principle component analysis is a multivariate statistical method which attempts to describe the total variation in a multivariate sample with fewer variables than in the original data set suggested by Johnson and Wichern (2003). A principle component depend solely on the covariance matrix $\sum$ (or the correlation matrix ρ)

Table 1. Average annual rainfall of 42 mandals in Visakhapatnam district (1986-2010).

S.No	STATION	MEAN (mm)	CV (%)	S.No	STATION	MEAN (mm)	CV (%)
1	Chinthapalli (CHNTP)	1308.7	33.3	22	Rambilli (RMBL)	885.4	30.8
2	Koyyuru (KYR)	1228.2	23.6	23	Chodavaram (CHVR)	1013.6	25.5
3	G.K.Veedhi (GKVD)	1262.5	31.7	24	Ravikamatham (RVKM)	936.5	34.3
4	Paderu (PDR)	1125.1	18.8	25	Butchiapeta (BTCH)	869.7	29.6
5	G.Madugula (GMDL)	1274.5	25.6	26	Anakapaalli (AKP)	951.3	24.6
6	Munchingput (MNCP)	1413.4	31.8	27	Kasimkota (KSMK)	880.4	25.9
7	Pedabayalu (PDBY)	1154.9	24.5	28	K.Kotapadu (KKTP)	999.5	20.0
8	Hukumpeta (HKMP)	1082.7	18.7	29	Sabbavaram (SBVM)	988.0	27.8
9	Araku Vally (ARKV)	1132.4	17.1	30	Paravada (PRVD)	848.2	23.3
10	Ananthagiri (ANTG)	1173.1	27.5	31	Visakhapatnam (VSKP)	910.7	31.5
11	Dumbriguda (DMBG)	1129.2	26.8	32	Pendurthi (PNDR)	859.9	28.5
12	Madugula (MDL)	1075.7	32.0	33	Bheemunipatnam (BMNP)	885.4	30.7
13	Narsipatnam (NRSP)	1041.9	24.2	34	Padmanabam (PDNB)	1002.3	42.2
14	Golugonda (GLGN)	1036.4	27.0	35	Anandapuram (ANDP)	968.0	25.2
15	Rolugunta (RLGN)	974.5	29.5	36	Munagapaka (MNGP)	825.7	30.8
16	Kotauratla (KTRT)	901.3	30.2	37	Gajuwaka (GJWK)	880.9	28.1
17	Makavarapalem(MKVRP)	1004.4	24.0	38	Pedagantyyada (PDGNT)	811.0	28.5
18	Nathavaram (NTVR)	1024.5	27.6	39	Payakarao peta (PRPT)	954.9	29.2
19	Nakkapalli (NKP)	975.0	34.8	40	S.Rayavaram (SRVR)	883.8	33.1
20	Elamanchili (ELMN)	957.9	26.7	41	Cheedikada (CDKD)	803.4	21.9
21	Atchutapuram (ATCP)	873.3	25.3	42	Devarapalli (DVP)	981.1	26.9
	MEAN						**1006.9**
	SD						**144.8**

of $X_1, X_2, \ldots\ldots\ldots, X_p$ through a few linear combinations of the original variables. The eigen value and eigen vector pairs created from data matrix is utilized to identify the principle components.

Let the random vector $X=[X_1, X_2, \ldots, X_p]$ have the covariance matrix Σ with eigen values $\lambda_1 \geq \lambda_2 \geq \ldots. \geq \lambda_p \geq 0$. Then the linear combinations $Y_1=\lambda_1'X$, $Y_2=\lambda_2'X$,...., $Y_p=\lambda_p'X$.

The first principle component is the linear combination with the maximum variance i.e. it maximizes Var $(Y_1) = \lambda_1'\Sigma\lambda_1$. It is clear that Var $(Y_1) = \lambda_1'\Sigma\lambda_1$ can be increased by multiplying any λ_1 by some constant. To eliminate this indeterminacy, it is convenient to restrict attention to coefficient vector of unit length.

Then, Var $(Y_i) = \lambda_i'\Sigma\lambda_i$ i=1, 2, ……,p (1)

Cov $(Y_i, Y_k) = \lambda_i'\Sigma\lambda_k$ i, k=1, 2, ……,p

The principle components are those uncorrelated linear combinations $Y_1, Y_2, \ldots\ldots, Y_p$ whose variances in (1) are larger as possible.

The eigen values and eigen vectors were computed from data matrix. Eigen values define the amount of total variation that is displayed on principle components. The proportion of variation accounted for each principle component (PC) is explained as the eigen value divided by the sum of eigen values.

$$\boxed{\text{Proportion of total population variance due to pth principle component}} = \frac{\lambda_p}{\lambda_1 + \lambda_2 + \lambda_3 + \ldots\ldots + \lambda_p} \quad (p = 1, 2, \ldots..., k)$$

$$= \frac{\lambda_p}{\text{Trace}(S)}$$

The eigen vector (loadings) defines the correlation of each variable with the principle

components. The correlation between the k^{th} original variable X_k and the i^{th} principle component Y_i is given by

$$\rho_{y_i,x_k} = \frac{l_{ki}\sqrt{\lambda_2}}{\sqrt{\sigma_\kappa}} \qquad i, k = 1, 2, \ldots, p$$

Here σ_κ is the standard deviation of and (λ_1, ℓ_1), (λ_2, ℓ_2),……, (λ_p, l_p) are the eigen value-eigen vector pairs for Σ. Rainfall patterns were obtained for all the 42 mandals of rainfall data covering years 1986-2010. The rainfall data were analyzed by using the statistical software, SAS.

RESULTS AND DISCUSSION

The Visakhapatnam district shows considerable year to year variability in the rainfall recorded over the years. Considering this variability in the rainfall, the classification was obtained in eight clusters. This classification provided, as far as possible, homogenous groups of mndals with similar rainfall. The mandal wise average annual rainfall received during the period of study i.e., 1986 to 2010 is presented in the Table 1.

Munchingput mandal was identified as the highest rainfall receiving mandals with an average annual rainfall of 1413.4 mm. It was observed that the lowest average annual rainfall of 803.4 mm was identified in Cheedikada mandal. The variability of rainfall over the years was found to be of high order in Padmanabam 42.2 per cent (Table 1). This variability in the rainfall among the mandals thus forms a basis for the classification.

Correlation matrix was used for principle component analysis and the results of the principle component analysis are presented in the Table 2. It was observed that the eigen values, which represent the measure of similarity between the Common Principal Component (CPC) and vector subspaces. The first component eigen value is considerably high (λ_1=10.68) and the remaining component eigen values are almost similar (λ_2=2.36, λ_3=1.60,λ_4=1.45, λ_5=1.29 and λ_6=1.15). The results thus indicate that all the mandals are close together along the first six CPCs.

The first principle component contributed maximum towards variability (42.74%) and second and third principle components described 9.47 per cent and 6.41 per cent respectively. The vector sub space of first CPC is loaded on Year 2003 (0.267), while rainfall of years 1992 (0.330) and 1999 (0.606) are heavily loaded in second and third CPCs. It was observed that the first six principle components with eigen values more than one are considered as statistically significant and contributed 74.22 per cent of cumulative variability among 42 mandals of rainfall.

The rainfall pattern in Visakhapatnam district can be divided into three regions based on deviation from the mean annual rainfall of Visakhapatnam district (Table 1). These three regions are high rainfall regions (>Mean+SD), median rainfall regions (Mean±SD) and low rainfall regions (<Mean-SD) was presented in Table 3.

The CPCs identified above formed the basis for clustering. The component scores corresponding to the mandals were then obtained on the basis of the mean vectors of the district. For the purpose of clustering, agglomerative hierarchical clustering method i.e., Ward's method was followed to group the entries into different clusters based on Euclidean distance. The ward's method of clustering can be applied for classification due to its several advantages over other procedures (Seber, 1984).

It was observed that majority of the mandals, i.e., 12 out of 42 mandals of data exhibited a similar pattern in the rainfall (Table 4). These mandals formed in clusters I and III, where as there were two single observation clusters. The mandals classified in these two single observation clusters were MNCP (VII) and PDNB (VIII) with an average annual rainfall of 1413.4 mm and 1002.3 mm respectively. The intra and inter cluster Euclidean distance values are displayed in figure 1. The intra cluster distance ranged from 0.00 to 49.3. Maximum intra cluster distance was found in cluster VI (49.3). The average inter cluster distance between clusters I and VII were maximum (169.7) and minimum (29.8) between clusters III and IV.

The mean cluster rainfall in principle component analysis over the study period (1986-2010) was given in the Table 4. It was observed that high rainfall was identified in the clusters V (1210.7 mm), VI (1285.6 mm) and VII (1413.4 mm). The remaining clusters were identified as medium rainfall clusters. No low rainfall was

Table 2. Common principle components for rainfall data of Visakhapatnam district.

Year	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6
1986	0.148	0.203	-0.229	0.455	-0.033	0.064
1987	0.096	0.168	-0.323	-0.297	0.425	-0.267
1988	0.138	0.155	-0.199	0.158	0.492	0.285
1989	0.189	0.304	-0.135	0.118	0.214	0.110
1990	0.237	0.066	0.055	-0.203	0.066	0.309
1991	0.217	0.130	0.256	-0.146	-0.142	0.069
1992	0.140	0.330	0.027	0.340	-0.051	0.049
1993	0.171	0.246	-0.037	-0.067	-0.342	0.086
1994	0.115	-0.146	0.198	-0.209	0.086	0.578
1995	0.223	0.223	0.142	-0.275	-0.087	-0.074
1996	0.186	0.313	0.185	-0.227	-0.212	0.076
1997	0.242	0.163	-0.104	-0.112	0.037	-0.181
1998	0.175	0.134	0.013	-0.142	-0.025	-0.420
1999	0.015	0.127	0.606	0.247	0.256	-0.127
2000	0.179	-0.215	0.379	0.195	0.235	-0.112
2001	0.217	-0.046	0.163	0.201	-0.050	-0.275
2002	0.230	-0.188	-0.063	0.180	-0.076	0.106
2003	0.267	-0.177	0.015	0.032	-0.047	0.026
2004	0.254	-0.040	-0.113	0.184	-0.143	0.084
2005	0.234	-0.246	-0.157	0.054	-0.168	0.061
2006	0.205	-0.155	-0.120	0.137	-0.222	-0.131
2007	0.247	-0.238	-0.107	-0.017	-0.023	-0.097
2008	0.226	-0.061	0.003	-0.122	0.078	-0.054
2009	0.231	-0.175	0.031	-0.059	0.262	-0.058
2010	0.198	-0.313	-0.038	-0.169	0.125	-0.048
Eigen value	10.68	2.36	1.60	1.45	1.29	1.15
% of variance explained	42.74	9.47	6.41	5.83	5.16	4.62

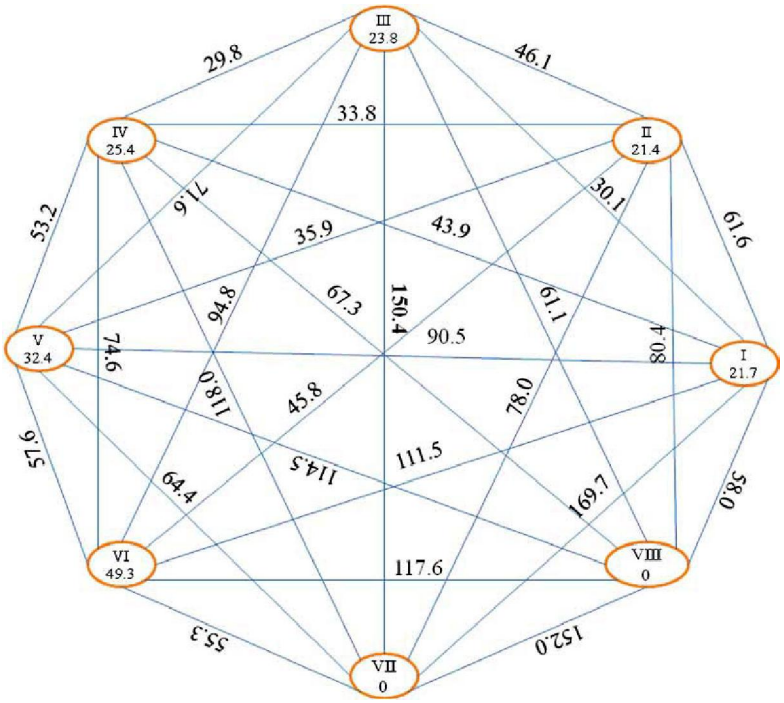
Table 3. Rainfall pattern based on deviation from mean annual rainfall of Visakhapatnam district.

S.No	Category	Formula	Rainfall (mm)
1	High rainfall	>Mean+SD	>1162
2	Medium rainfall	Mean±SD	862-1162
3	Low rainfall	<Mean-SD	<862

Table 4. Clustering pattern and Mean cluster rainfall (mm) in principle component analysis.

Cluster	I	II	III	IV	V	VI	VII	VIII
Mandals	PRVD, PDGNT, CDKD, VSKP, BTCH, MNGP ATCP, GJWK, PNDR, DVP, PRPT, SRVR,	PDBY, PDR, HKMP, ARKV, ANTG	ELMN, NKP, KTRT, ANDP, SBVM, RMBL AKP, KKTP, BMNP, RVKM, KSMK, RLGN	MDL, GLGN, MKVPM, CHVR, NRSP, NTVRM	KYR, GMDL, DMBG	CHNTP, GKVD	MNCP	PDNB
No. of mandals	12	5	12	6	3	2	1	1
Mean	875.2	1133.7	942.0	1032.8	1210.7	1285.6	1413.4	1002.3

Fig.1. Configuration of clusters and their mutual relationship by Ward's minimum variance method using PCA scores (Not to scale)



identified in these clusters. The mandals that were found in different clusters exhibits similar rainfall patterns over the study period.

CONCLUSIONS

The study confined to the 25 years mandalwise rainfall data of Visakhapatnam district of Andhra Pradesh. The rainfall analysis in Visakhapatnam district indicated that multivariate approaches based on principle component analysis effectively summarized the source of variability in the rainfall and precisely identifies the different rainfall patterns, which are likely to occur in different mandals of the district. The application of this approaches identified the principle components more than one were statistically significant, and contributed 74.22 per cent of rainfall variability and identified medium rainfall (862 mm-1162 mm) was the most frequent representative pattern of rainfall in majority mandals of Visakhapatnam district. The approaches can also be applied for identifying the rainfall patterns, which would help in planning suitable crop strategies on the basis of the availability of rainfall during the crop period.

LITERATURE CITED

Hills R C and Morgan J H 1981 Rainfall Statistics: An interactive approach to analysis of rainfall records for agricultural planning. *Experimental Agriculture,* 17: 1-16.

Kulkarni B S and Pandit N S N 1988 A discrete step in the technology for sorghum yields in Parbhani, India. *Agricultural and Forest Meteorology,* 42: 157-165

Kulkarni B S and Rao N G 2000 The common principle component approach for clustering under multiple sampling. *Journal of Indian Society of Agricultural Statistics*, 53: 1-11

Kulkarni B S and Reddy D 1994 The cluster analysis approach for classification of Andhra Pradesh on the basis of rainfall. *Mousam,* 45: 325-332.

Prabhakaran P V, Pillai P B and Karamchandran K M 1992 Climatic classification of Kerala. *Mousam*, 43: 434-436.

Seber G A F 1984 *Multivariate Observations*. John Wiley and Sons, New York, USA, Pp 359-366.

Johnson R A and Wichern D W 2003 *Applied Multivariate Statistical Analysis*. 3rd ed., Prentice Hall of India, New Delhi, Pp 356-370.

(Received on 20.06.2013 and revised on 10.01.2014)