Mysore J. Agric, Sci., 46 (2) : 293-299, 2012

Cluster Analysis on Area, Production and Productivity for Major Selected Crops in Karnataka

SANTOSHA RATHOD, H. S. SURENDRA, R. MUNIRAJAPPA AND H. CHANDRASHEKAR
Department of Agricultural Statistics, Applied Mathematics and Computer Science,
College of Agriculture, UAS, GKVK, Bangalore-560 065

ABSTRACT

The study was carried out to make Clusters based on area, production and productivity for major selected crops in Karnataka. Data regarding different crops considered for the study were obtained from various issues of 'Karnataka at a Glance' for the period of 1985-2005. The study spans over a period of 20 years from1985 to 2005 and was divided into two phases (i) Pre WTO (World Trade Organization) era from 1985-1995 and (ii) Post WTO era from 1995-2005. Ward's method of clustering has been used in this study to make clusters based on area, production and productivity. The Dendrogram shows the graphical representation of the results of hierarchial cluster analysis. The study revealed that, for first period (1985-95) cropped area of sorghum, cotton, paddy, groundnut and ragi showed similarity in area. For second period (1995-2005) paddy, maize, mango, ragi and groundnut had similarity in area. Sugarcane and sorghum had similarity in production for first period and sugarcane alone for second period. Based on productivity major clusters were formed by horticulture crops for both periods.

THE cluster analysis is a multivariate analysis which deals with grouping of objects or individuals into unknown objects based on the homogeneity of the objects. This method involves an agglomerative clustering algorithm. Ward's method starts out with *'n'* clusters of size one and continues until all the observations are included into one cluster.

The Dendrogram shows the graphical representation of the results of hierarchial cluster analysis. There was a tree-like plot where each step of hierarchial clustering was represented as a fusion of two branches of the tree into a single one. The branches represent clusters obtained on each step of hierarchial clustering.

MATERIAL AND METHODS

To perform cluster analysis cultivated area, production and productivity following crops were considered *viz.,* paddy, ragi, bajra, wheat, jowar, maize, bengalgram, redgram, groundnut, sugarcane and cotton. The horticulture crops considered are mango, citrus, banana, guava, grapes and papaya for the period of 1985-2005 (KAG-1985-2005). The data of study spans over a period of 20 years from1985 to 2005 and was divided into two phases: (i) Pre WTO (World Trade Organization) era from 1985-1995 and (ii) Post WTO era from 1995-2005.

Cluster analysis usually done to combine cases into groups when the group membership was not known prior to the analysis. Cluster analysis is a technique for grouping individual or objects into unknown groups. The classification proceeds from the most general to the most specific in steps. The most general classification was kingdom followed by phylum, sub-phylum and class etc. (Richard A Johnson and Dean W .Wichern, 2002).

Clustering methods: The commonly used methods of clustering fall into two general categories. (i) Hierarchial (ii) Non hierarchial. Hierarchial clustering techniques proceed by either a series of mergers or a series of successive divisions.

Euclidean distance: Euclidean distance indicates the most commonly used measures of association (similarity) that are frequently used with continuous variables in cluster analysis. For instance, in two dimensions, one can plot the observations in a scatter diagram, and measure the distances between the pairs

of points. Generally the following equation can be used :

$$d(\mathbf{X}_i, \mathbf{X}_j) = \sqrt{\sum_{k=1}^{p} (X_{ik} - X_{jk})^2}$$

X_{ik} = Response for variable *k* in sample unit *i* (the number of individual species *k* at site *i*)

X_{jk} = Response for variable *k* in sample unit *j* (the number of individual species *k* at site *j*)

n = Number of sample units

p = Number of variables

For each variable *k*, take the difference between the observations for sites *i* and *j*. These differences are then squared, and summed over *p* variables. This gives us the sum of the squared difference between the measurements for each variable. Finally, take the square-root of the result.

Ward's method: It is an alternative approach for performing cluster analysis. Basically, it looks at cluster analysis as an analysis of variance problem, instead of using distance metrics or measures of association.

This method is most appropriate technique for analyzing quantitative variables, and not for binary variables.

Further, under this method the following concepts are defined:

Error Sum of Square (ESS) :

$$ESS = \sum_{i} \sum_{j} \sum_{k} (X_{ijk} - \bar{X}_{ik})^2$$

X_{ijk} denote the value for variable '*k*' in observation '*j*' belonging to cluster '*i*'.

Summing over all variables and all of the units within each cluster and comparing the individual observations for each variable against the cluster means for that variable. If the Error Sum of Square is small, then it suggests that data are closer to their cluster means, implying that a cluster of like units.

Total Sum of Squares (TSS):

$$TSS = \sum_{i} \sum_{j} \sum_{k} (X_{ijk} - \bar{X}_{..k})^2$$

The total sum of squares is defined in the similar way of ESS. Further, comparing the individual observations for each variable against the grand mean.

R-square (R^2) :

$$R^2 = \frac{TSS - ESS}{TSS}$$

The R^2 value is interpreted as the proportion of variation explained by a particular clustering of the observations.

Dendrogram : The Dendrogram is a graphical representation of the results of hierarchical cluster analysis. This appears in the form of tree like plot where, each step of hierarchical clustering is represented as a fusion of two branches of the tree resolving into a single one. The branches represent clusters obtained on each step of hierarchical clustering at 10 per cent semi partial R^2.

Results and Discussion

Eigen values of the Correlation Matrix for crop area during 1985-1995 : The results of the Eigen values of the Correlation matrix for the area of 18 different agricultural crops which are predominantly grown in the district for the period 1985-1995 are presented in Table I.

The table shows that the Eigen value and their corresponding proportion of variation as well as their cumulative proportion for 20 initial crop clusters. It can be seen that first, second and third, initial clusters have the Eigen values 13.450, 3.280, and 1.354, respectively which are greater than one, hence the Ward's method of clustering resulting only with three final clusters for this period of study. Further, the result depicts that first 17 districts resulting with cumulative value to one, during the period 1985-1995.

Crop cluster based on area during 1985-1995 : The cluster analysis was done based on the cultivated area of different crops across all the districts of Karnataka for the period of 1985-1995. From the

TABLE I

Eigen values of the Correlation Matrix for crop area during 1985-1995

Districts	Eigen values of the correlation matrix			
	Eigen value	Difference	Proportion	Cumulative
1	13.450	6.214	0.6725	0.6725
2	3.280	2.056	0.1640	0.8367
3	1.354	1.012	0.0677	0.9042
4	0.838	0.615	0.0419	0.9461
5	0.514	0.407	0.0257	0.9718
6	0.144	0.102	0.0072	0.9790
7	0.110	0.099	0.0055	0.9845
8	0.078	0.062	0.0039	0.9884
9	0.060	0.049	0.0030	0.9914
10	0.050	0.350	0.0025	0.9939
11	0.040	0.028	0.0020	0.9959
12	0.030	0.020	0.0015	0.9969
13	0.016	0.015	0.0001	0.9979
14	0.014	0.011	0.0008	0.9987
15	0.010	0.010	0.0008	0.9994
16	0.007	0.007	0.0005	0.9999
17	0.002		0.0001	1.0000

findings, it can be seen that only three final major clusters were observed.

The first cluster formed by cotton, paddy, ground nut and ragi. This indicates only these crops have similarity in area of cultivation across all the districts of Karnataka.

The second cluster was mainly formed with sorghum, this indicates that this crop has similarity in area of cultivation across all the districts of Karnataka. Because the sorghum was both food and fodder crop, so usually the crop is being grown in all the districts extensively for many purpose, that is why this crop formed separate cluster.

Third cluster was formed with remaining all other crops. This indicates that similarity in area of cultivation across all the districts of Karnataka. Based on the suitability of weather parameters, and soil conditions, etc. these crops are growing in all the districts. The overall performance of these crops across the Karnataka was similar.

The findings are in accordance with the results as reported by Mo Huidong (1987); Cali Ski and Harabasz (1992).

Eigen values of the Correlation Matrix for Crop area during 1995-2005 : The results are presented in Table II of the Eigen values from the Correlation matrix for the area of 18 different agricultural crops grown in the districts for the period of 1995-2005. The table shows that, the Eigen value and their corresponding proportion of variation as well as their cumulative proportion for 27 initial clusters. It can be seen that first, second and third, initial clusters have the Eigen values 20.612, 3.969 and 1.569, respectively which are having Eigen values greater than one. Hence, the Ward's method of clustering resulting only with three final clusters for period of study 1995-2005. However, the result depicts that first 15 districts resulting with cumulative value to one, during the period 1985-1995.

Crop cluster based on area during 1995-2005 : The cluster analysis was done based on the area of cultivation of different crops across all the districts of

TABLE II

Eigen values of the Correlation Matrix for crop area during 1995-2005

Districts	Eigen values of the correlation matrix			
	Eigen value	Difference	Proportion	Cumulative
1	20.612	8.162	0.7634	0.7634
2	3.969	0.093	0.1470	0.9804
3	1.569	0.714	0.0570	0.9674
4	0.400	0.126	0.0148	0.9822
5	0.154	0.098	0.0077	0.9899
6	0.143	0.064	0.0053	0.9952
7	0.046	0.021	0.0017	0.9962
8	0.027	0.014	0.0010	0.9979
9	0.016	0.009	0.0006	0.9985
10	0.014	0.008	0.0005	0.9990
11	0.011	0.007	0.0004	0.9994
12	0.008	0.004	0.0003	0.9996
13	0.005	0.003	0.0002	0.9998
14	0.003	0.002	0.0001	0.9999
15	0.002		0.00001	1.0000

Karnataka for the period 1995-2005. From the result, it is evident that only three major clusters were formed.

The first cluster was formed with paddy, maize and mango. Because paddy was the staple food crops of Karnataka, there was a commercialization in cropping system so mango and maize have major area in this particular period. Weather parameters are approximately similar for growing these crops with similar in area of cultivation practices.

The second cluster was mainly formed with ragi and groundnut due to the shift in cropping pattern, groundnut was mainly cultivated in this period, and ragi was also grown majorly, so usually this crop is being grown in all the districts extensively for many purpose.

Third cluster was formed by remaining all other crops. The overall performance of these crops across the Karnataka was similar. Third cluster was formed with remaining all other crops.

The findings are in accordance with the results as reported by Mohuidong (1987); Cali Ski and Harabasz (1992).

Eigen values of the Correlation matrix for crop production during 1985-1995 : The results of the Eigen values of the Correlation matrix for the production for the period of 1985-1995 are presented in Table III. The table shows that, the Eigen value and their corresponding proportion of variation as well as their cumulative proportion for 20 initial clusters. It was observed that first, second and third, initial clusters have the Eigen values 15.932, 2.386 and 1.178 greater than one, hence in the Ward's method of clustering only three final clusters are possible for this period of study. However, the findings indicate that first 12 districts alone resulting with cumulative value to one, for the period 1985-1995.

Crop cluster based on production during 1985-1995 : The cluster analysis done based on the production of different crops across all the districts of Karnataka for the period 1985-1995.

The first cluster was formed by sugarcane. Most of the districts have been growing this crop as major

TABLE III

Eigen values of the Correlation Matrix for crop production during 1985-1995

Districts	Eigen values of the correlation matrix			
	Eigen value	Difference	Proportion	Cumulative
1	15.932	6.509	0.7966	0.7966
2	2.386	1.107	0.1193	0.9159
3	1.178	0.921	0.0589	0.9748
4	0.214	0.321	0.0107	0.9855
5	0.102	0.103	0.0051	0.9906
6	0.076	0.098	0.0038	0.9944
7	0.048	0.092	0.0054	0.9970
8	0.026	0.012	0.0013	0.9983
9	0.014	0.009	0.0007	0.9990
10	0.012	0.008	0.0006	0.9996
11	0.006	0.002	0.0003	0.9999
12	0.002		0.0001	1.0000

crops in this period. Because of commercialization of this crop dominated in this particular period, the bio-mass of the crop was higher and also long duration crop.

The second cluster was mainly formed with sorghum. This was an important food and fodder crop across almost all the districts of Karnataka. Based on the suitability of weather parameters, soil conditions, etc. this crop is being grown in all the districts. The overall performance of this crop across the Karnataka was similar.

Third cluster was formed with remaining all other crops. This indicates that similarity in area of cultivation across all the districts of Karnataka. Based on the suitability of weather parameters, and soil conditions, etc. these crops are growing in all the districts. The overall performance of these crops across the Karnataka was similar.

Eigen values of the Correlation Matrix for crop production during 1995-2005 : The results of the Eigen values of the Correlation matrix for the production for the period of 1985-1995 are presented in Table IV. The table shows that, the Eigen value and

TABLE IV

Eigen values of the Correlation Matrix for crop production during 1995-2005

Districts	Eigen values of the correlation matrix			
	Eigen value	Difference	Proportion	Cumulative
1	23.158	7.105	0.8577	0.8577
2	3.086	2.019	0.1143	0.9720
3	0.594	0.408	0.0220	0.9940
4	0.405	0.216	0.0150	0.9955
5	0.024	0.012	0.0009	0.9964
6	0.022	0.010	0.0008	0.9972
7	0.019	0.008	0.0007	0.9979
8	0.016	0.007	0.0006	0.9985
9	0.014	0.001	0.0005	0.9990
10	0.011	0.001	0.0004	0.9994
11	0.008	0.000	0.0003	0.9997
12	0.005	0.003	0.0002	0.9998
13	0.003	0.002	0.0001	0.9999
14	0.003		0.0001	1.0000

their corresponding proportion of variation as well as their cumulative proportion for 18 initial clusters. It was observed that first and second clusters have the Eigen values 23.158 and 3.086, respectively which are greater than one. Hence in the Ward's method of clustering only two final clusters are possible for this particular period of study. However, the results established that first 14 districts alone resulting with cumulative value to one, for the period 1995-2005.

Crop cluster based on production during 1995-2005 : The cluster analysis was done based on the production of different crops across all the districts of Karnataka for the period 1995-2005. Here two final clusters were formed.

The first cluster was formed by the sugarcane alone. This was an important commercial crop across almost all the districts of Karnataka. The bio-mass of the crop was higher and it was a long duration crop. Based on the suitability of weather parameters, soil conditions, etc these crops are being grown in all the districts. The second cluster was mainly formed with other remaining crops. The overall performance of these crops across the Karnataka was almost similar.

The reported results of this study goes with the findings and conclusions of Hardeman *et al.* (2003); Fleming *et al.* (2004).

Eigen values of the Correlation Matrix for crop productivity during 1985-1995 : The results of the Eigen values of the Correlation matrix for the productivity are presented in Table V. The table shows that, the Eigen value and their corresponding proportion of variation as well as their cumulative proportion for 20 initial clusters.

From the findings it can be observed that first, second and third, initial clusters have the Eigen values 15.000, 3.174 and 1.936, respectively which are greater than one, hence in Ward's method of clustering resulting with only three final clusters for period of study (1985-1995). However, the results established that that first 14 districts alone resulting with cumulative value to one, for the period 1985-1995.

Crop cluster based on productivity during 1985-1995 : The cluster analysis done based on the productivity of different crops across all the districts of Karnataka for the period of 1985-1995 .Three final clusters formed.

TABLE V

Eigen values of the Correlation Matrix for crop productivity during 1985-1995

Districts	Eigen values of the correlation matrix			
	Eigen value	Difference	Proportion	Cumulative
1	15.000	5.216	0.7500	0.7500
2	3.174	1.091	0.1587	0.9087
3	1.936	0.742	0.0968	0.9755
4	0.290	0.149	0.0145	0.9900
5	0.066	0.106	0.0033	0.9933
6	0.048	0.096	0.0024	0.9957
7	0.032	0.058	0.0016	0.9973
8	0.020	0.041	0.0010	0.9983
9	0.012	0.013	0.0006	0.9999
10	0.008	0.009	0.0004	0.9993
11	0.006	0.007	0.0003	0.9996
12	0.004	0.005	0.0002	0.9998
13	0.002	0.002	0.0001	0.9999
14	0.002		0.0001	1.0000

The first cluster was formed by papaya. This is notified as an important horticulture crop across almost all the districts in Karnataka. The second cluster was mainly formed with sapota, mango, grapes, sugarcane and banana and third cluster was formed by all other remaining crops. Based on the suitability of weather parameters, and soil conditions, etc. these crops are being grown in all the districts, the overall performance of these crops across the Karnataka was similar.

By looking into the clusters formed based on productivity major clusters were formed by the horticulture crops even though they are grown in less area but productivity of these crops are high.

Eigen values of the Correlation Matrix for crop productivity during 1995-2005 : The results of the Eigen values of the Correlation matrix for the productivity for the period 1995-2005 are presented in Table VI. The table shows that, the Eigen value and their corresponding proportion of variation as well as their cumulative proportion for 27 initial clusters.

From the table it can be concluded that first, second and third, initial clusters have the Eigen values 18.994, 4.957 and 2.195, respectively which are greater than one. Hence in Ward's method of clustering only three final clusters are possible for this period of study. Further, the findings reveal that that first 13 districts alone resulting with cumulative value to one, for the period 1995-2005.

Cluster based crop productivity during 1995-2005 : The cluster analysis done based on the productivity of different crops across all the districts of Karnataka for the period 1985-1995. Three final clusters were formed.

The first cluster was formed by the crops papaya and sugarcane. The second cluster was mainly formed by banana. In the second period also horticulture crops such as papaya and banana dominated in the production, and the sugarcane notified an important commercial crop in most of the districts of Karnataka. The bio-mass of the crop was higher and it was long duration crop. The third cluster was mainly formed with remaining all other crops. Based on the suitability of weather parameters, and soil conditions, etc. these crops are growing in all the districts, the overall performance of these crops across the Karnataka was similar.

TABLE VI

Eigen values of the Correlation Matrix for crop productivity during 1995-2005

Districts	Eigen values of the correlation matrix			
	Eigen value	Difference	Proportion	Cumulative
1	18.994	5.136	0.7035	0.7035
2	4.957	1.509	0.1836	0.8872
3	2.195	0.813	0.0812	0.9684
4	0.427	0.210	0.0158	0.9853
5	0.154	0.096	0.0057	0.9910
6	0.095	0.035	0.0035	0.9945
7	0.070	0.031	0.0026	0.9971
8	0.027	0.026	0.0010	0.9981
9	0.024	0.020	0.0009	0.9990
10	0.014	0.010	0.0004	0.9994
11	0.008	0.006	0.0003	0.9997
12	0.005	0.003	0.0002	0.9999
13	0.003		0.0001	1.0000

The reported results of this study found on par with the findings and conclusions of Kenlenbeck and Maas (2005).

Overall performance of the cluster analysis : For first period, area of crops between 1985-1995, various crops such as sorghum, cotton, paddy, groundnut and ragi showed similarity in area across all the districts in Karnataka. For the second period of 1995-2005, crops like paddy, maize, mango, ragi and groundnut had similarity in area across all the districts of Karnataka. When the performance of crop clusters based on area between two periods was compared, it was evident that cotton has shifted by introduction of other crops like maize and mango over the years in the second period of study (Table VII).

For the period of 1985-1995, sugarcane and sorghum had similarity in production across all the districts of Karnataka. For the period of 1995-2005, only sugarcane showed similarity in production across

TABLE VII

Overall performance of the cluster analysis

Aspects	Study period (year)	First cluster	Second cluster	Third cluster
Area	1985-1995	Cotton, paddy, groundnut and ragi (4 crops)	Sorghum(1 crop)	Other remaining crops (13 crops)
	1995-2005	Paddy, maize and mango (3 crops)	Ragi and Groundnut (2 crops)	Other remaining crops (12 crops)
Production	1985-1995	Sugarcane (1 crop)	Sorghum (1 crop)	Other remaining crops (16 crops)
	1995-2005	Sugarcane (1 crop)	Other remaining crops (17 crops)	–
Productivity	1985-1995	Papaya (1 crop)	Sapota, mango, sugarcane, Banana and grapes (5 crops)	Other reamining crops (12 crops)
	1995-2005	Papaya and sugarcane (2 crops)	Banana (1 crop)	Other remaining crops (15 crops)

all the districts of Karnataka. When comparison of the performances of crop clusters based on production between two periods was observed, it indicated that sorghum production shifted over the period (Table VII).

For the first period (1985-1995) papaya, sapota, mango, sugarcane, banana and grapes had similarity in productivity across all the districts of Karnataka. For second period 1995-2005, the crops like papaya and sugarcane showed similarity in productivity across all the districts of Karnataka. When the performances of crop clusters were compared based on productivity between two periods, it indicated that mango and banana production are shifted over the years (Table VII).

REFERENCES

CALI SKI, T. AND HARABASZ, J., 1992, Dendrite Method for Cluster Analysis, *Communications in Statistics - Simulation and Computation*, **3**(1) : 1–27.

FLEMING,. L., WESTFALL, D. G., WIENS, D.W. AND BRODAHL. M. C., 2004, Evaluating Farmer Defined Management Zone Maps for Variable Rate Fertilizer Application, *Precision Agriculture*, **2** (2): 201-215.

HARDEMAN, R. T., 2003, Use of cluster analysis for identification and classification of farming systems in Qing Yang County, Central North *China, Agron. J.,* **95:** 584-596.

KARNATAKA AT A GLANCE (KAG)., Directorate of Economics and Statistics- 1985-2005.

KEHLENBECK, K. AND MAASS., B. L., 2005, Crop Diversity and Classification of Home gardens in Central Sulawesi: Indonesia, *Agro forestry Systems,* **63**(1): 53-62.

MO HUIDONG GU SHILIANG, 1987, Cluster Analysis for Agronomic Characters of Barley Varieties in Jiangsu-Zhejiang Shanghai Area Scientia Agriculture Sinica, **30** (7), 143-157.

RICHARD, A. JOHNSON. AND DEAN W. WICHERN., 2002, *Applied Multivariate Statistical Analysis,* PHI Publication New delhi, *V Ed*, pp 668-748.

(*Received :* August, 2011 *Accepted :* February, 2012)