International Journal of Modern Agriculture, Volume 10, No.2, 2021
ISSN: 2305-7246

Clustering Of Villages Based On Soil Parameters

Dr.A. Rajarathinam[1] and M. Ramji[2]
[1]Professor, Department of Statistics, Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli, India-627012. E-mail: arrathinam@yahoo.com
[2]Reasearch Scholar, Department of Statistics, Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli, India-627012. E-mail: 7jipgm@gmail.com

Abstract

An empirical investigation was carried out to study the variation in five soil data, including potential of Hydrogen, Electrical Conductivity, Organic Carbon, available Phosphorus, and Potassium. The data on these five soil parameters pertaining to 47 villages of Palayamkottai taluka in the district of Tirunelveli, Tamil Nadu State, INDIA were obtained from the Soil Health Card scheme. These soil parameters were subjected to various statistical analyses. An analysis of variance showed that variations in different soil parameters among the villages were highly significant; that is, these individual parameters were significantly different across the villages. A multivariate analysis of variance test revealed a significant variation between the villages when all the five soil parameters were considered simultaneously. Through all the soil parameters were found to be significant both individually and together, the clustered variation was largely due to variations in Organic Carbon, Electrical Conductivity, and Phosphorus as confirmed by Ward's method. Three clusters were identified such that there was homogeneity within the clusters and heterogeneity between the clusters.

Keywords: Analysis of variance, Multivariate analysis of variance, Squared Euclidean Distance, Ward's Method, Fusion coefficient

1. INTRODUCTION

The soil health data bank which was recently developed by the Tamilnadu state agricultural department, INDIA as part of the soil health card (SHC) programme is one the voluminous set of data. By considering variability in these data, various suitable statistical techniques can be deriving possible conclusions. The five soil nutrient content, including Potential of Hydrogen (pH), electrical conductivity (EC), organic carbon (OC), Phosphorus (P), and Potassium (K) are the parameters most analyzed by soil-testing laboratories across the country. An analysis of OC, as well as available P and K, help to determine the recommended rate of N, P, and K fertilizers. EC and pH guide as to the choice of crops and the soil management practices that can be used to enhance soil productivity.

The specific objectives of the study are as follows:

(i) To estimate the hidden variability pattern in the soils of the villages in Palayamkottai taluk.
(ii) To test the significance of variation in each soil nutrient content between the villages in Palayamkottai taluk using Analysis of variance (ANOVA).
(iii) To study the variability pattern of all the five soil parameters considered simultaneously between the villages of Palayamkottai Taluk using Multivariate Analysis of Variance (MANOVA).
(iv) To group villages into clusters based on the variability patterns of the soil parameters using Ward's clustering method and to study the characteristics of these clusters.
(v) To establish the geographical closeness among the villages based on soil parameters.

2. MATERIALS AND METHODS

An attempt was made to group the 47 villages of Palayamkottai Taluk based on five soil parameters namely EC (ds/m); OC (%); P (P_2O_5 kg/ha); K (K_2O in kg/hg), using Ward's method. In each village, 20 soil samples are collected.

2.1. Analysis of Variance

An ANOVA technique (Rao 1952) is employed to test the significance of the variation in each parameter between the villages. We use the following model,

$$y_{ij} = \mu + v_i + e_{ij} \qquad (1)$$

i= 1, 2, 3... 47; j=1, 2, 3 … 20

where y_{ij} is the status of the soil parameter in the jth villages of the ith sample, μ is average status, v_i is the status in the ith village, and e_{ij} are a random error which follows a normal distribution with mean zero and constant variance σ^2.

2.2. Multivariate Analysis of Variance

To test the significance of variation among all the five parameters considered simultaneously a MANOVA technique (Johnson and Wichern 2002) is employed. The MANOVA model for comparing the population mean vectors (g=47) is as follows.

$$Y_{ij} = \mu + V_i + E_{ij} \qquad (2)$$

i= 1, 2, 3... 47; j=1, 2, 3 … 20

where E_{ij} is a vector of random errors distributed as $N_p(0, \Sigma)$ (p=1, 2, 3, 4, 5). Here the parameter vector μ is the overall mean and V_i represents the status of the model in (2), each component of the observation vector Y_{ij} satisfies the univariate model (1) and the variance-covariance matrix Σ is the same for all populations.

2.3 Cluster Analysis

To address within-variability of the village mean values, different soil nutrient contents mean values are converted into uncorrelated variables using the pivotal condensation method (Rao 1952). The transformed uncorrelated variables are used to group the villages with Ward's method which involves squared Euclidean distance method. The optimum number of clusters is calculated based on cluster selection criteria enumerated in Aldenderfer and Blashfield (1984)

2.4 Inter and Intra Cluster distance

After the formation of the clusters, inter and intra-cluster D^2 values are calculated using the averaged individual D^2 values. The square root of this D^2 is used to indicate inter and intra-cluster distances. The cluster means for all characters are computed using the character means for the villages included in the clusters.

2.5 Estimation of intra and inter-cluster variance for different characters

An un-weighted analysis of variance using the mean values of different characters is implemented (Rao, 1952). The structure of the analysis of variance is given below.

Variation type	Degrees of freedom	Mean squares	Expected mean squares
Between Cluster	(k-1)	MSB (say)	$\sigma_w^2 = m\sigma_b^2$
Within Cluster	$\sum_{i=1} n_i - k$	MSW (say)	σ_w^2

MSB= Mean square between the clusters; MSW= Mean square within-cluster; k= number of clusters; n_i = number of villages in the i^{th} cluster; m is the harmonic mean based on number of villages in each cluster.

Using the mean squares, the estimates of inter and intra-cluster variances (i.e. σ_w^2 and σ_b^2) are obtained for each cluster. Also, the ratio of the inter-cluster variances to the total variances is obtained as follows.

$$R^2 = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2}$$

The inter-cluster co-efficient of variations is calculated as follows.

$$CV_b = \frac{\sqrt{\sigma_b}}{\overline{X}} \times 100$$

Note that $\overline{X}$ is the general mean for the character.

3. RESULTS AND DISCUSSION

The results are presented below on statistical analysis of data on five soil nutrient contents from 20 soil samples taken in each of the 47 villages of Palayamkottai Taluka.

3.1 Analysis of variance

The ANOVA results of the data presented in Table 1 reveals that for each of the soil nutrient contents, the mean square values and F-test results were significant, and this result reveals that there exists a variation in individual parameters between the villages.

Table 1. Analysis of variance of different soil parameters

Type of variation	Degrees of freedom	Mean Squares				
		pH	EC	OC	P	K
Between villages	46	1.936	0.020	0.052	284448.813	229756.897
Within villages	912	0.304	0.029	0.009	6510.783	7340.555

The pH values of Palayamkottai Taluk varied from 5.73 and 8.46, with a mean of 7.489 (Table 6). The minimum mean value for the village of Thimmarajapuram was at par with that of Palayamkottai-II. The mean values for the remainder of the villages were significantly different from that of Thimmarajapuram. Most of these values were similar to the maximum mean values of Karaieruppu.

The minimum, maximum, and mean values of EC were 0.069 ds/m, 0.889 ds/m, and 0.395 ds/m, respectively (Table 6). A high variation in mean values was observed between the villages with significant differences. The maximum mean value of Nochikulam was distinctly higher than that of the other villages.

The OC value of Palayamkottai Taluk varied from 0.16% to 0.69% with a mean of 0.36% (Table 6). The maximum mean value of 0.692 was distinctly higher than that of Thiruvannanathapuram, Therkupatti, and Palayamkottai-II. The minimum mean value of Kattarankulam was at par with of the mean values of Karuppandurai, Kurichikulam, and Melaputhaneri.

Regarding P, the minimum, maximum, and average values were 11.88 kg/ha to 535.00 kg/ha and 199.97 kg/ha, respectively. The maximum mean values of Thiruvannanathapuram were at par with the mean values of Keelaveeraragavapuram, Palayamkottai-II, and Chathiram Pudukulam (Table 6).

The K values of Palayamkottai Taluk varied from 133.00 kg/ha to 737.63 kg/ha, with a mean of 390.471 kg/ha. The minimum value of Kunnathur was at par with the mean values of Keelapattam, and Kuruchilkulam. The reminders of the village's mean values were distinctly different from that of Kunathur (Table 6).

3.2 Multivariate Analysis of Variance

Different multivariate tests, including Pilla’s trace, Wilk’s lambda, Hotelling's trace, and Roy's largest root tests were employed for testing the joint variation of all five soil nutrient contents across the villages. The results are presented in Table 2.

A review of the statistical Multivariate Analysis of Variance results presented in Table 2 shows that the p-values of different multivariate test statistics were equal to 0.000 indicating highly significant differences between the village-level mean values when all five parameters are considered simultaneously. This implies that when considered together, the five nutrient contents showed heterogeneous values across the villages.

Table 2. Characteristics of MANOVA statistics

Effect	Multivariate Tests	Value of test statistics	Value of F	Significance
Parameters, EC, pH, OC, P, and K	Pilla's Trace	1.299	15.202	.000
	Wilk's Lambda	0.110	16.077	.000
	Hotelling's Trace	4.348	16.956	.000
	Roy's Largest Root	3.182	26.089	.000

3.3 Mahalanobis' D^2

The minimum value of D^2 (0.000) was observed between the villages of Chathiram pudukulam and Sivandipatti. the maximum value of D^2 (431.361) was observed between the villages of Anadha Krishnapuram and Vellakoil.

3.4 Cluster Analysis

To address within-variability of the village mean values, different soil parameter means values given in Table 6 are converted into uncorrelated variables using the pivotal condensation method described in Rao (1952). The transformed uncorrelated variables are used to group the villages.

3.5 Ward's Method

The Ward's method Johnson and Wichern (2002) with squared Euclidean Distance was used to group the villages. The Dendrogram generated by the Ward's method depicted the number of clusters. The optimum cluster selection procedure described in Aldenderfer and Blashfield (1984) was used to find the number of clusters. As the fusion co-efficient value was suddenly changed in the third cluster, as per the Fig.1., a three-cluster pattern was identified. The dendrogram produced by Ward's method is depicted in Fig.2. The distributions of villages in different clusters are depicted in Fig.3. Also the distribution of 47 villages in different three clusters and the cluster means for the five soil nutrient contents are presented in Table 3.

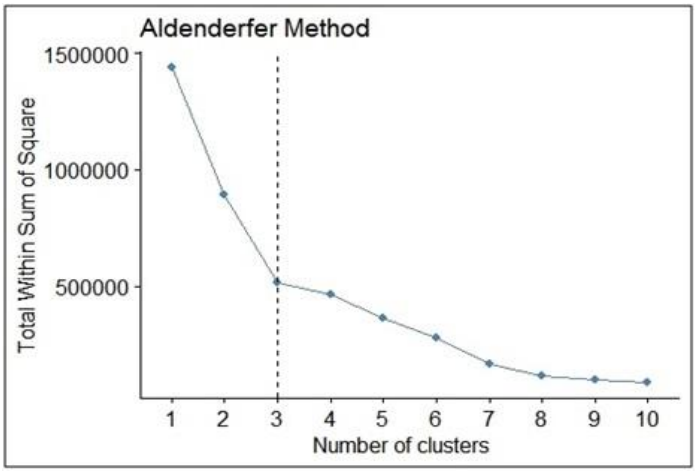

Fig.1. Optimal number of clusters

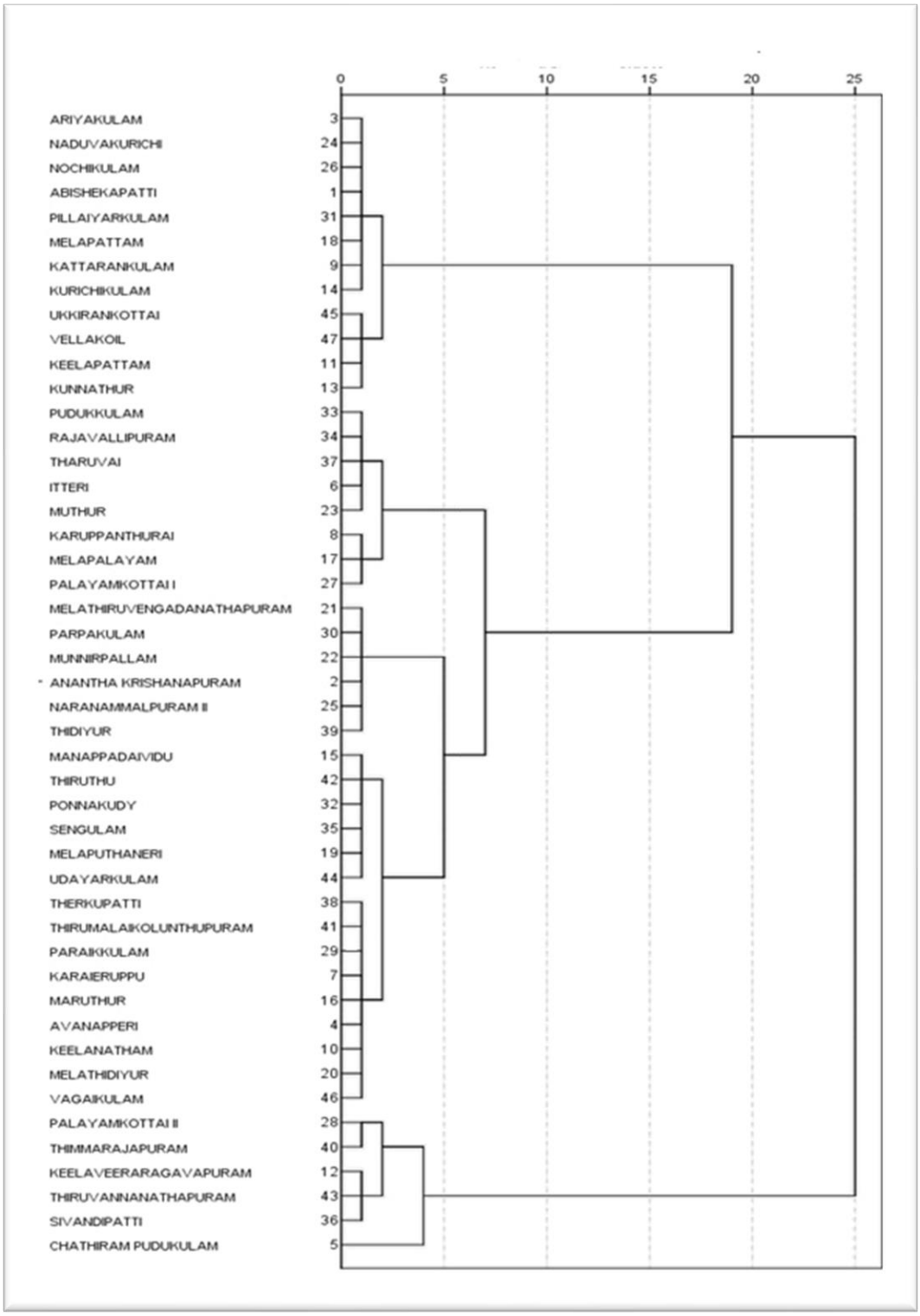

Fig.2. Ward's minimum variance dendrogram formed by the villages of Palayamkottai taluka.

Cluster 2 contains maximum number of 29 villages followed by 12 in Cluster I and 6 in Cluster III. The clustering pattern revealed that geographic diversity may not be necessarily related to village diversity; rather it may be due to soil heterogeneous patterns.

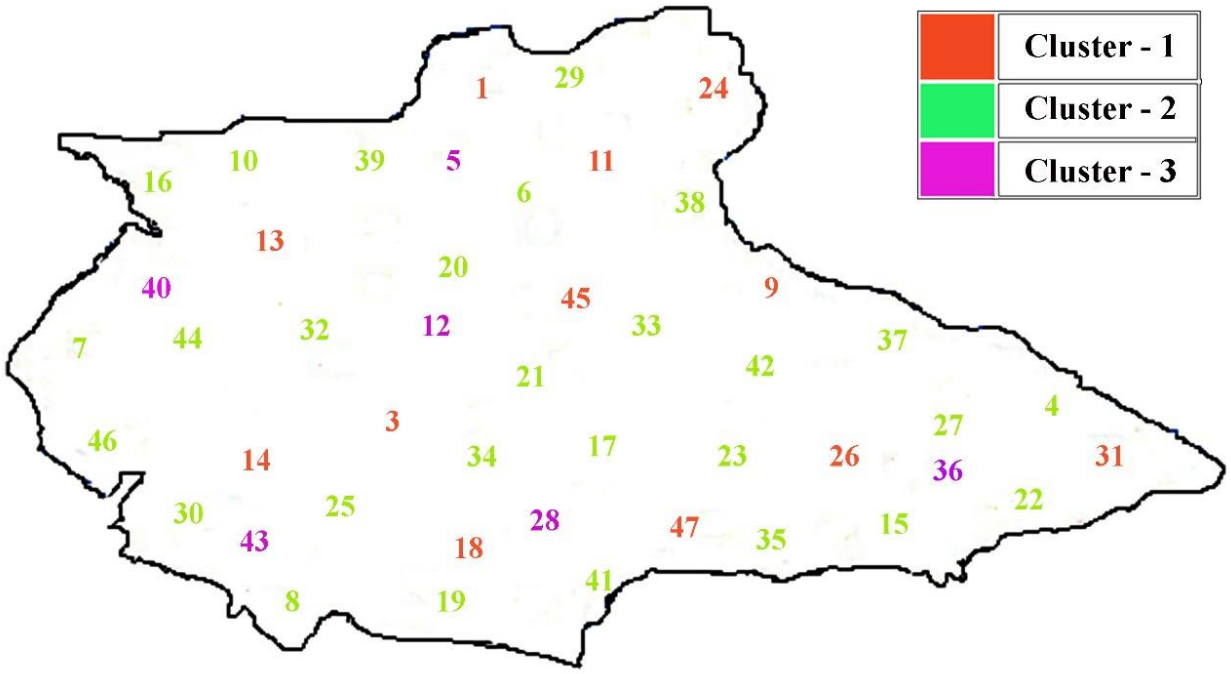

Fig.3. Distribution of villages in different clusters of Palayamkottai taluka.

Among the mean values of pH for different clusters, the value of 6.73833 of cluster III was the lowest value among the clusters, whereas the value of 7.60879 of cluster II was the maximum. The difference between the maximum and minimum mean value is 0.87046. This indicated that the variation is very low.

The EC mean values for the different clusters varied from 0.39750 ds/m (cluster I) to 0.46800 ds/m (cluster III). Very negligible variations were observed; see Table 3. Regarding the mean values of OC, the minimum value was observed for cluster I at 0.28200% whereas the maximum was observed for cluster III at 0.42367%.

For the mean value of P, cluster I showed the lowest value at 138.11042 kg/ha whereas the maximum mean value of 486.12917 kg/ha was observed for cluster III. The mean values for K varied from 221.15625 kg/ha (cluster I) to 457.68750 kg/ha (cluster III).

Table 3. Cluster composition and mean values based on Ward's method.

Cluster No.	No.of Villages	Villages / Mean Values					
I	12	Village Names	Abishekapatti, Ariyakulam, Kattarankulam, Keelapattam, Kunnathur, Kurichikulam, Melapattam, Naduvakurichi, Nochikulam, Pillayarkulam, Ukkirankottai, Vellakoil.				
		Mean Values	pH	EC	OC	P	K
			7.5754	0.39750	0.28200	138.11042	221.15625
II	29	Village Names	Anantha Krishnapuram, Avanapperi, Itteri, Karaieruppu, Karuppanthurai, Keelanatham, Manappadaividu, Maruthur, Melapalayam, Munnirpallam, Muthur, Naranammalpuram-II, Melaputhaneri, Melathidiyur, Melathiruvengadanathapuram, Palayamkottai-I, Paraikkulam, Parpakulam, Ponnakudy, Pudukkulam, Rajavallipuram, Sengulam, Tharuvai, Therkupatti, Thidiyur, Thirumalaikolunthupuram, Thiruthu, Udayarkulam, Vagaikulam.				
		Mean Values	pH	EC	OC	P	K
			7.60879	0.37914	0.37590	166.36592	446.62466
III	6	Village Names	Chathiram Pudukulam, Keelaveeraragapuram, Palayamkottai-II, Sivandipatti, Thimmarajapuram, Thiruvannanathapuram.				
		Mean Values	pH	EC	OC	P	K
			6.73833	0.46800	0.42367	486.12917	457.68750

3.6 Inter and intra-cluster distance.

The results presented in Table 4 are Intra and inter-cluster distances of the clusters. Intra cluster distance measured in terms of D^2 values ranged from 7.634 in cluster I to 83.333 in cluster II. Meanwhile, inter-cluster distance in terms of D^2 values ranged from 83.333 in cluster II to 254.369 in cluster III.

Table 4. Mean intra-and inter-cluster D^2 values obtained using Ward's Method.

Custer	I	II	III
I	7.634*	0.0000	5.825
II		83.333*	0.00000
III			254.369*

*Diagonal values indicate the intra-cluster distance.

3.7 Inter and Intra-Cluster Variance

The analysis of variance for each of the five soil nutrient contents was carried out using means of the 47 villages in the three different clusters. To identify which parameter most determined the formation of the three clusters, two benchmarks were used, including R^2 (i.e. Ratio of inter-cluster variance to total variance) and Adjusted R^2 (i.e inter-cluster insignificant of the model). These values were estimated for each of the five soil nutrient contents. The data are presented in Table 5.

A maximum R^2 value of 0.66 was obtained for P, 0.59 for K, 0.22 for pH, 0.20 for OC, and 0.03 for EC. The result indicated that the formation of clusters based on these five parameters was largely due to variations in EC, OC, pH, K, and P values.

Table 5. The ratio of inter-cluster variance to the total variance (R^2) and intra-cluster co-efficient of variations (CV_b)

Sl. No	Parameters	R^2	CV_b (%)
1	**pH**	0.22	7.56
2	**EC**	0.03	43.32
3	**OC**	0.20	30.72
4	**P**	0.66	78.18
5	**K**	0.59	33.56

The maximum CV_b value was 78.18% for P, which was followed by 43.32% for EC, 33.56% for K, 30.72% for OC and 7.56% for pH. Thus, based on R^2 and co-efficient variations in P, K, and pH contributed most to cluster formation.

4. CONCLUSION

The variations in all five individual soil parameters among the 47 villages under study were highly significant, indicating that the selected variables significantly differed among the villages. The variability within villages was non-significant due to homogeneity within the villages. A MANOVA test revealed significant variability between the villages of the Palayamkottai Taluk when all the five soil nutrient contents were considered simultaneously. Cluster formation was largely determined by the variations in pH, EC, and OC values, as confirmed by Ward's method. These grouping can be to prepare fertility maps and to develop effective soil improvement programs. Though distinct clusters were identified in Palayamkottai Taluk, Geographical closeness among the villages in each cluster was not evident, indicating that the factors influencing the soil parameters were geographically well dispersed.

Table. 6 Mean values of soil parameters for villages in Palayamkottai Taluk

Sr.No.	Name of the Villages	pH	EC	OC	P	K
1	ABISHEKAPATTI	6.235	0.527	0.372	127.375	287.125
2	ANANTHA KRISHANAPURAM	7.445	0.437	0.362	135.250	586.500
3	ARIYAKULAM	7.950	0.239	0.310	74.500	235.875
4	AVANAPPERI	7.140	0.539	0.357	153.125	422.500
5	CHATHIRAM PUDUKULAM	6.785	0.477	0.229	489.125	737.625
6	ITTERI	8.105	0.443	0.360	73.875	543.875
7	KARAIERUPPU	8.460	0.223	0.334	166.500	353.375
8	KARUPPANTHURAI	7.135	0.154	0.202	41.500	417.250
9	KATTARANKULAM	7.330	0.795	0.155	118.250	188.500
10	KEELANATHAM	7.950	0.557	0.259	122.125	423.500
11	KEELAPATTAM	7.520	0.487	0.277	255.125	164.375
12	KEELAVEERARAGAVAPURAM	7.385	0.512	0.315	505.500	442.625

13	KUNNATHUR	6.645	0.147	0.290	184.250	133.000
14	KURICHIKULAM	7.280	0.483	0.174	126.475	167.125
15	MANAPPADAIVIDU	7.645	0.471	0.342	317.500	354.000
16	MARUTHUR	7.010	0.293	0.437	169.875	325.000
17	MELAPALAYAM	8.035	0.324	0.271	11.875	407.750
18	MELAPATTAM	7.915	0.229	0.323	138.750	234.625
19	MELAPUTHANERI	7.135	0.475	0.223	257.000	432.750
20	MELATHIDIYUR	7.645	0.503	0.563	167.375	471.750
21	MELATHIRUVENGADANATHAPURAM	7.360	0.589	0.304	222.000	540.370
22	MUNNIRPALLAM	7.780	0.522	0.388	271.875	537.750
23	MUTHUR	8.045	0.387	0.453	26.250	534.375
24	NADUVAKURICHI	7.895	0.206	0.316	76.875	228.750
25	NARANAMMALPURAM II	7.825	0.158	0.362	186.950	587.250
26	NOCHIKULAM	7.525	0.889	0.274	55.500	234.250
27	PALAYAMKOTTAI I	6.995	0.416	0.454	48.638	330.250
28	PALAYAMKOTTAI II	5.795	0.345	0.543	494.750	308.750
29	PARAIKKULAM	7.755	0.407	0.361	212.750	370.000
30	PARPAKULAM	7.730	0.448	0.425	220.125	531.620
31	PILLAIYARKULAM	8.185	0.221	0.276	82.125	276.125
32	PONNAKUDY	7.445	0.324	0.496	325.500	424.500
33	PUDUKKULAM	8.055	0.220	0.353	44.250	477.625
34	RAJAVALLIPURAM	7.935	0.497	0.331	60.250	468.000
35	SENGULAM	6.740	0.437	0.380	296.500	441.375
36	SIVANDIPATTI	7.450	0.627	0.402	467.500	497.375
37	THARUVAI	7.970	0.600	0.277	55.875	499.875
38	THERKUPATTI	8.065	0.365	0.596	192.125	350.250
39	THIDIYUR	7.285	0.429	0.446	126.250	511.500
40	THIMMARAJAPURAM	5.730	0.312	0.363	424.000	293.125
41	THIRUMALAIKOLUNTHUPURAM	7.710	0.375	0.472	214.375	344.250
42	THIRUTHU	6.470	0.215	0.279	284.900	366.000
43	THIRUVANNANATHAPURAM	7.285	0.538	0.692	535.900	466.625
44	UDAYARKULAM	8.030	0.128	0.415	235.250	466.500
45	UKKIRANKOTTAI	8.410	0.300	0.292	189.750	266.500
46	VAGAIKULAM	7.755	0.069	0.409	184.750	432.375
47	VELLAKOIL	8.015	0.253	0.327	228.350	237.625
	Mean	7.489149	0.395304	0.357979	199.9726	390.4705
	Minimum	5.73	0.0685	0.1545	11.875	133
	Maximum	8.46	0.8885	0.692	535.9	737.625
	Standard Deviation	0.617794	0.170868	0.107242	136.8817	131.0359
	Co-efficient of variation (%)	8.249182	43.22442	29.9577	68.4502	33.55846

pH- potential of Hydrogen, **EC**-Electrical Conductivity, **OC**- Organic Carbon, **P**- Phosphorus, **K**- Potassium

ACKNOWLEDGEMENT

The authors are thankful to the referee for providing valuable suggestions to improve the quality of the paper.

References

[1] M.S.Aldenderfer, R.K.Blashfield, Cluster Analysis, Sage Publications, Newbury Park, California, 1984.

[2] R.A.Johnson, W.D.Wichern, Applied Multivariate Statistical Analysis, Prentice-Hall of India Pvt.Ltd, New Delhi, 2002.

[3] C.R.Rao, Advanced Statistical Methods in Biometrical Research. John Wiley and Sons, New York, USA, 1952.